

Alfred Inselberg

Foreword by Ben Shneiderman

Parallel Coordinates

Visual Multidimensional Geometry and Its Applications



 Springer



CD-ROM

Parallel Coordinates

Visual Multidimensional Geometry
and Its Applications

To

Hadassah, Dona and Haim, Avigail, Ofra, Yuval, Akiva
... a wonderful family

and

Bernie Dismdale, Lew Leeburg, Alex Hurwitz,
Misha Boz, Baxter Armstrong, Ralph Alterowitz
... great friends and colleagues

and

Heinz von Foerster
... Mischiefologist, Magician,
Cybernetician, and Teacher

Parallel Coordinates

Visual Multidimensional Geometry
and Its Applications

with 230 color illustrations

Alfred Inselberg

Foreword by Ben Shneiderman



CD-ROM



Springer

Alfred Inselberg
Tel Aviv University
School of Mathematical Sciences
69 978 Tel Aviv
Israel
aiisreal@post.tau.ac.il

ISBN 978-0-387-21507-5 e-ISBN 978-0-387-68628-8
DOI 10.1007/978-0-387-68628-8
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2008927731

Mathematics Subject Classification (2000): 15-xx, 51-xx, 62-xx, 62-09, 68-xx, 14J26, 14J29, 14J70: 15-01, 15A04, 15A06; 51-01; 51A05, 54-01; 68-01, 68P01, 68W01; 90-01, 90-08, 90C05, 90C29, 90C59

© Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The breakthrough idea of parallel coordinates has enchanted many people with its cleverness and power. Al Inselberg, the source of that cleverness and power, finally shares the depth and breadth of his invention in this historically important book.

I've been waiting impatiently for this exposition since I was captivated by Inselberg's lecture at the University of Maryland in 1979. He had already understood the potential parallel coordinates has for generating insights into high-dimensional analysis and information visualization. In the following decades he polished the arguments, built effective software, and demonstrated value in important applications. Now a broad community of readers can benefit from his insights and effective presentation.

I believe that Inselberg's parallel coordinates is a transformational ideal that matches the importance of René Descartes' (1596–1650) invention of Cartesian coordinates. Just as Cartesian coordinates help us understand 2D and 3D geometry, parallel coordinates offer fresh ways of thinking about and proving theorems in higher-dimensional geometries. At the same time they will lead to more powerful tools for solving practical problems in a wide variety of applications. It is rare to encounter such a mind-shattering idea with such historic importance.

While Inselberg's insight and exposition opens the door to many discoveries, there is much work to be done for generations of mathematicians, computer scientists, programmers, and domain experts who will need to build on these innovative ideas.





To understand and apply parallel coordinates, many further innovations are needed. Those who can readily reshape their perceptual skills and realign their cognitive frames will be well-equipped for this nontrivial task. I still struggle with these novelties, which Inselberg has masterfully accomplished. In each of my dozen meetings with Inselberg, he relieved me of my confusion and steered me closer to clarity. There is a chance that those who learn parallel coordinates early in life may be able to more easily see higher-dimensional spaces. More likely, master teachers, inspired interface designers, and gifted domain experts will enable future generations to grasp these profound ideas as if they were entirely natural, just as we see Cartesian coordinates.










I encourage readers to read each paragraph carefully and make sure they fully grasp the ideas. Inselberg has made this possible by his lucid and often charming












prose, filled with intriguing historical references and clever quips. The computer scientists, programmers, dataminers, and information visualization experts may want to jump ahead to Chapter 10 which has practical examples and numerous screen shots. The recent additional insights from Inselberg's students and followers give a hint of the thousands of papers that are likely to be inspired by this work. More importantly, the applications of parallel coordinates could contribute to solving some of the complex multidimensional problems of our time. I hope every reader will strive to put these ideas to work in making a better world.



















Ben Shneiderman
University of Maryland, College Park
May 2008



















Contents

Foreword	v
Preface	xv
Acknowledgments	xxiii
1 Introduction	1
1.1 Multidimensional Visualization	1
1.2  <i>FT-1</i> How?	2
2 Geometry Background	7
2.1 Preliminaries	7
2.1.1 In the Beginning	7
2.2  <i>FT-1</i> Why Projective Geometry?	9
2.2.1 Projective Plane Model	10
2.2.2 Axioms for Projective Geometry	13
2.2.3 Principle of Duality	16
2.3 ** Finite Geometries	19
2.3.1 Seven-Point Geometry	19
2.3.2 Finite Geometry with 13 Points	21
2.4  <i>FT-2</i> Analytic Projective Geometry	22
 <i>FT-2e</i>	24
2.4.1 Homogeneous Coordinates on a Line	25
2.5 The Fundamental Transformations of Projective Geometry	26
2.6 More on Homogeneous Coordinates	29
2.6.1 Linearizability of Projective Transformations	29
2.7 ** The Linearized Projective Transformations	31
2.7.1 Projective Transformations in the Plane \mathbb{P}^2	32
2.7.2 Projective Transformations in \mathbb{P}^3	33
2.8 ** Coordinate System Rotations	36
2.9 ** Analytic Proofs	39

2.9.1	Computational Proof	39
2.9.2	Algebraic Proof	41
2.10	** Conics on the Projective Plane	43
2.10.1	In the Projective Plane All Conics Are Ellipses	43
2.10.2	Generalizing Pappus's Theorem	44
2.11	** Geometries and Invariants	45
3	 FT-1 The Plane with Parallel Coordinates	49
3.1	The Fundamental Duality	49
3.2	Transformations under the Duality	57
3.2.1	Rotations and Translations	57
3.2.2	** Recognizing Orthogonality	58
3.3	A Preview	61
4	Multidimensional Lines	63
4.1	 FT-1 Representing Lines in \mathbb{R}^N	63
4.1.1	Elementary Approach	63
	 FT-1e	68
4.1.2	 FT-2 Some Properties of the Indexed Points	70
	 FT-2e	73
4.1.3	Representation Mapping I	73
4.1.4	** The General Case	75
4.1.5	** Construction Algorithms	79
4.1.6	Convenient Display of Multidimensional Lines	83
4.1.7	 FT-3 Rotations and Translations	87
4.2	Distance and Proximity Properties	88
4.2.1	Intersecting Lines	88
4.2.2	Nonintersections	89
	 FT-3e	91
4.2.3	Minimum Distance between Two Lines in \mathbb{R}^N	91
4.2.4	** Air Traffic Control	100
5	Planes, p -Flats, and Hyperplanes	115
5.1	 FT-1 Planes in \mathbb{R}^3	115
5.1.1	Vertical Line Representation	115
5.1.2	** Planar Coordinates	118
	 FT-1e	124

5.2	 FT-2	Representation by Indexed Points	124
5.2.1		The family of “Superplanes” \mathcal{E}	125
5.2.2		The Triply Indexed Points	127
	 FT-2e	135
5.3	 FT-3	Construction Algorithms	136
5.3.1		Planes and Lines	136
5.3.2		The Four Indexed Points	138
	 FT-3e	141
5.3.3		Special Planes	141
	 FT-4	144
5.3.4		Intersecting a Plane with a Line	145
	 FT-4e	146
5.3.5		Points and Planes: Separation in \mathbb{R}^3	146
5.3.6		** Separation in \mathbb{R}^3 : An Old Approach	149
5.3.7	 FT-5	Rotation of a Plane about a Line and the Dual Translation	151
5.4		Hyperplanes and p -Flats in \mathbb{R}^N	159
5.4.1		The Higher-Dimensional Superplanes	159
5.4.2		Indexed Points in \mathbb{R}^N	164
	 FT-6	168
	 FT-6e	168
5.4.3		Collinearity Property	169
	 FT-7	170
	 FT-7e	170
5.5		** Construction Algorithms in \mathbb{R}^4	174
5.5.1		The Five Indexed Points	174
5.5.2		Intersecting Hyperplanes in \mathbb{R}^4	175
5.6		Detecting Near Coplanarity	179
5.7		Representation Mapping, Version II	180
6	** Envelopes		185
6.1		The Basic Idea	185
6.2		Formulation	186
6.3		Necessary and Sufficient Conditions	188
6.3.1		Singular Points	189
6.4		Examples: Envelopes of Families of Curves	190

7	Curves	195
7.1	 FT-1 Point-Curves and Line-Curves	195
7.1.1	Separation in the xy Plane	196
7.2	Duality between Cusps and Inflection Points	199
	 FT-2	200
7.3	 FT-3 Point-Curves \rightarrow Point-Curves	204
7.4	 FT-4 Curve Plotting	207
7.4.1	Space-Curves	213
7.5	 FT-5 Transforms of Conics	216
7.5.1	** Proving the Duality Conics \leftrightarrow Conics	217
7.5.2	 FT-6 Classification of the Conic Transforms	220
	 FT-6e	225
7.5.3	** The Conic Transform's Ideal Points	226
7.6	 FT-7 Transforms of Algebraic Curves	229
7.7	 FT-8 Convex Sets and Their Relatives	231
7.7.1	Gconics and Their Transforms	232
	 FT-8e	232
7.7.2	** Operational Dualities with Gconics	238
8	Proximity of Lines, Planes, and Flats	243
8.1	 FT-1 Motivation and a Topology for Proximity	243
8.2	Proximity of Lines and Line Neighborhoods	244
	 FT-1e	250
8.3	 FT-2 Proximity of Hyperplanes	251
8.3.1	Formulation of the Problem in \mathbb{R}^N	251
8.3.2	Outline of the Development	253
	 FT-2e	254
8.3.3	Properties of f_N	254
8.3.4	The Region Ω	257
	 FT-3	259
8.3.5	** Construction of Ω	263
	 FT-4	264
	 FT-4e	275
8.3.6	** The Full Image \overline{NH}	275
	 FT-5	283

8.3.7	Some Examples	283
 FT-6	285
8.3.8	Matching Points and Navigation within \overline{NH} . . .	287
9	Hypersurfaces in \mathbb{R}^N	297
9.1	 FT-1 Preliminaries	297
9.2	Formulation	301
9.3	 FT-2 Boundary Contours	304
	 FT-2e	306
9.4	 FT-3 Developable Surfaces	310
	 FT-3e	314
9.4.1	 FT-4 Ambiguities and Uniqueness	316
	 FT-4e	317
9.4.2	Reconstruction of Developable Surfaces	320
9.4.3	Specific Developables	325
	 FT-5	326
	 FT-6	333
	 FT-7	336
	 FT-7e	336
9.4.4	 FT-8 General Developable Surfaces	342
9.4.5	** Higher Dimensions	344
9.5	 FT-9 Ruled Surfaces	346
	 FT-9e	348
9.6	** Approximate Quadric Hypersurfaces	359
9.6.1	The Hyperellipsoids	360
9.6.2	The Hyperellipse	362
9.6.3	Intelligent Process Control	363
9.7	 FT-10 More General Surfaces	365
	 FT-10e	372
9.7.1	** A More General Matching Algorithm	373
10	 FT Data Mining and Other Applications	379
10.1	Introduction	379
10.1.1	Origins	380
10.1.2	The Case for Visualization	380

10.2	Exploratory Data Analysis with $\ $ -coords	382
10.2.1	Multidimensional Detective	382
10.2.2	An Easy Case Study: Satellite Data	383
10.2.3	Quasars Dataset	390
10.2.4	Compound Queries: Financial Data	392
10.2.5	Hundreds of Variables	400
10.2.6	Production of VLSI (Chips)	401
10.3	Classification	406
10.3.1	Case Study: Noise Signature Recognition	412
10.4	Visual and Computational Models	416
10.5	Parallel Coordinates: The Bare Essentials	418
10.5.1	Lines	418
10.5.2	Planes and Hyperplanes	419
10.5.3	Nonlinear Multivariate Relations: Hypersurfaces	423
10.6	Future	425
11	Recent Results	429
11.1	Displaying Several Lines Efficiently	429
11.1.1	Abstract	429
11.1.2	Introduction	429
11.1.3	General Approach	431
11.1.4	Flipping Axes	432
11.1.5	Rotating Axes	435
11.1.6	Transforming Axes	438
11.1.7	Back to Rotating Axes	441
11.2	Separating Point Clusters on Different Planes	443
11.2.1	Overview	443
11.2.2	The Algorithm	443
11.2.3	Mean and Variance Estimation	445
11.2.4	Error Probability Approximation	448
11.2.5	Complexity	450
11.3	Surface Representation Decomposition and Developable Quadrics	451
11.3.1	The Hyperplane Representation as a Linear Transfor- mation on the Coefficients	452
11.3.2	A Brief Review of Duality in Projective Space	452
11.3.3	Extending the Projective Duality to Degenerate Quadrics	453

11.3.4	The Decomposition and Its Application to Developable Quadrics	456
11.3.5	A Numerical Example: An Elliptical Cone in 5-Space	457
11.4	Network Visualization and Analysis with Parallel Coordinates	460
11.4.1	Introduction	460
11.4.2	The NEVIS Transformation	461
11.4.3	Properties of NEVIS	464
11.4.4	Large-Scale Networks	469
11.5	To See \mathbb{C}^2	476
11.5.1	The Complex Field \mathbb{C}	476
11.5.2	The Space \mathbb{C}^2	477
11.5.3	Representing Points on a Line in \mathbb{C}^2	479
11.5.4	Holomorphic Functions	483
12	Solutions to Selected Exercises	489
12.1	Solutions to Exercises in Chapter 2	489
12.2	Solutions to Exercises in Chapter 3	509
12.3	Solutions to Exercises in Chapter 4	514
12.4	Solutions to Exercises in Chapter 5	520
	Notation and List of Symbols	531
	Bibliography	535
	Index	547

This page intentionally left blank

Preface

What and Why

In the late 1980s, researchers in computer graphics recognized a specific stream of applications as an emerging field and called it *visualization*. By now it has become a discipline with its own journals, conferences, and community of active workers applying it to science, city planning, entertainment, and much more. Even our reality is becoming more virtual. Yet since time immemorial people have expressed themselves visually, realizing in due course insight can also be *derived* from images [63]. In 1854, a cholera epidemic raging in a London neighborhood prompted Dr. John Snow to search for remedies. He examined the data consisting of a table with the addresses of those who had died from the disease. On a map of the neighborhood, which fortunately had the positions of the water wells, he placed dots at the locations of the recorded deaths. The concentration of dots in the vicinity of just one well was *visually* striking. He became convinced that there was a connection, and had the handle of the suspect well replaced. The epidemic stopped! Apparently the disease was being transmitted by contact with the handle. This true story is widely considered an early success of visualization [172]; the picture provided insight no one had gleaned from the table of the data.

Legend has it that Archimedes was absorbed in a diagram when a Roman soldier killed him; the first recorded death in defense of visualization! “Do not disturb my circles” he pleaded as he was being struck by the sword. Visualization flourished in geometry, where visual interaction with diagrams is interwoven in the testing of conjectures and construction of proofs. Our tremendous ability in pattern recognition enables us to extract insight from images. This essence of visualization is abstracted and adapted in the more general problem-solving process to the extent that we form a mental image of a problem we are trying to solve and at times we say *see* when we mean understand.

My interest in visualization was sparked and nourished while learning Euclidean geometry. Later, while studying multidimensional geometries I became frustrated by the *absence* of visualization. Basically, we were working with equations, interpreting them geometrically, without the enjoyment and benefit of pictures. I kept wondering about ways to draw “pictures” of multidimensional objects. Of course, it would also be great fun doing *synthetic* high-dimensional constructions and have

tools for accurate visualization of multivariate problems. These thoughts, best attributed to the impetuosity of youth, absorbed me. What coalesced from this struggle is *parallel coordinates*.

A superficial resemblance to nomography [19] was noted early in the development [97]. “Nomograms are drawings made on the plane in order to replace cumbersome numerical calculations” ([144] p. 9), a technique that declined with the advent of modern computation. I was unaware until recently² of d’Ocagne’s marvelous monograph on *parallel coordinates* in two dimensions [141] with a *point* \leftrightarrow *line* correspondence. D’Ocagne was interested in the computational applications of nomography [142] rather than the development of a multidimensional coordinate system, which is where we come in. Invented independently, *parallel coordinates* is a *general-purpose VISUAL multidimensional coordinate system*. There are theorems on the unique representation of multidimensional objects, and geometrical algorithms for intersections, containment, minimal distances, proximities, etc., with a wide variety of applications.

The writing is undertaken with trepidation, knowing what kind of book it should not be: not a text in mathematics, though it should contain the foundations and development, nor computer science nor the other components of the methodology, nor a source-book for the variety of included applications. Rather it is an amalgam of all these, comprehensible and appealing to mathematicians, computer scientists, statisticians, engineers, scientists, and all those wanting to acquire deeper understanding of their multivariate problems by way of visualization.

In due course, it is hoped that the methodology will not only find its way into textbooks on various applications, but also be adopted for *visual experimentation and explanations* in teaching mathematics and the sciences at the university, high-school, and even elementary-school levels.

Style and Structure of the Book

This is a textbook with ample material for a one-semester course, easily adapted to shorter periods (see the following section), and containing advanced seminar topics. The writing evolved during a decade of teaching “visualization of multidimensional geometry and its applications,” a popular elective course in the applied mathematics/computer science curricula. It is also a reference for self-study or a companion text for courses on information visualization (particularly the chapter on data mining), visualization in general, data mining, GIS, management, statistics, linear algebra, analytic geometry, complex variables, and fields dealing with multivariate problems

²I am indebted to M. Friendly for pointing this out in [63].




(economics, engineering, physics, psychology, management, arts, and social sciences). The parallel coordinates methodology provides a multidisciplinary visual language and paradigm.

The mantra is to let intuition guide the formalism starting with the motivation for the emergence of parallel coordinates (\parallel -coords). Then the 2-dimensional plane is viewed with \parallel -coords as compared to Cartesian coordinates, leading to the discovery of surprising and pleasing properties. The chapters on the representation of lines, hyperplanes, curves, and topologies for proximity of lines and hyperplanes are interlaced with applications on collision-avoidance algorithms for air traffic control, geometric modeling, computer vision, statistics, and more. On this foundation the representation of surfaces is built, yielding exciting results such as viewing *convexity in any dimension*, recognizing nonorientability (as for the Möbius strip), and applications to intelligent instrumentation, process control, and decision support. There follows an *easy-to-read chapter on data mining and information visualization* with many fun applications (exploratory data analysis (EDA) on satellite, financial, vehicle recognition from noise-signature, astronomical and industrial datasets, automatic classification, decision support, nonlinear models). Results too recent to be incorporated in the text are included in the appendix at the end. They complement the representation of lines, developable surfaces, visualization and analysis of large networks, detecting and separating clusters of coplanar points, and the visualization of complex-valued functions: *to see \mathbb{C}^2* . The emphasis is on *visualization*, though computational and analytical aspects are treated rigorously, and where needed, extensively. Each chapter contains a variety of exercises, easy ones at first to check the understanding of the basics and more with increasing difficulty. Some exercises have served well as semester project topics. Where appropriate, open questions and research directions are pointed out. In places, details are deferred to references when technicalities may obscure the underlying ideas or unnecessarily intimidate the reader. I hope that the more mathematically sophisticated readers will forgive the occasional “sleights of hand.”

To make the book reasonably self-contained, after the introduction, there is a chapter on the needed geometrical background. It covers projective geometry, explaining the notions of *duality* and *homogeneous coordinates* as used in the subsequent rigorous development. The application to perspective, as in computer graphics, is developed with the realization that the “points at infinity” on the perspective plane are images of ideal points (directions). There is also a short intuitive chapter on the theory of envelopes; helpful background for the ensuing chapters on curves and surfaces. The attached CD contains the *interactive learning module* (ILM) for self-study and class demonstrations. There are pointers and

guidance for its use at the relevant sections in the text. Projective geometry theorems, dualities, transformations, and properties of representations are understood quickly and enjoyably.

FT Fast Track: How to Skip Reading Most of the Book

In each chapter the key portions are marked successively by  **FT-5**. The numeral, here 5, indicates the fifth fast-track marker within the chapter. The essentials can be swiftly grasped by following these road signs and returning for additional details later if needed. Typically the  **FT-5** symbol is placed at the beginning of a section to be read to the end or up to an “end marker” such as  **FT-5e**. The landmarks³ may also be figures. Covering these highlights is also a good way to review the material. The fast-track entries appear in the table of contents for sections or subsections. Portions marked with ** contain in-depth treatment of specific topics and are not needed for the basic ideas and results. Those interested in the applications can read the chapter on data mining and then scan the text for more details. Otherwise, choose your goal and follow the illuminated path:

- the essentials can be mastered quickly with the FT option;
- for extensive and sound foundations, include unmarked portions;
- to reach expert level, add some ** sections.

Course Organization

Much has been learned from the teaching experience. Reasonably well prepared and motivated students enjoyed and absorbed the material readily, contributing new results and ideas. After some experimentation, the course outline that proved most successful is the presentation of a course overview in the first two to three hours, including some demonstrations. This motivates the subject with problems and examples and also shows why, due to the 2-D point \leftrightarrow line duality, parallel coordinates need to be considered in the projective rather than the Euclidean plane. The development as given in the chapter sequence can then be safely followed. Demonstrations with the *ILM* in the CD as well as several applications are always well received. The instructor can organize group projects that provide such demonstrations. Some of the ones already given are available over the Internet. Also, there is an increasing collection of software commercially or freely available that is suitable for class demonstrations and use.

³I am indebted to Chao-Kuei Hung for this wonderful idea.

Instead of exams, the students, usually in groups of two or three, do projects. They are encouraged to choose topics in some area in which they have a personal interest and apply what they have learned in the course. Such projects have involved expanding the mathematical foundations, development and implementation of algorithms, writing programs particularly for data mining and modeling, and more. Examples of actual projects include “visualization of the Internet,” “collision detection and intersection in billiards,” “visualizing the ‘learning’ process in neural networks,” “topologies for proximity of flats (higher-dimensional planes),” several geometrical construction algorithms, “characterizing painters (i.e., impressionists, etc.),” “characterizing music,” data mining on a great variety of multivariate datasets (i.e., pollutants, bioinformatics, weather data, computer dating, constituents of various medicines and foods, etc.), differentiation and integration, fractals and linear programming with $\|$ -coords, large network visualization, visualizing complex-valued functions, and other topics. A wide variety of software on representations of surfaces, curve-plotters, implementation of results (i.e., translations \leftrightarrow rotations), etc. were produced. Some projects were subsequently presented at conferences, and others led to master’s theses. Student are encouraged to pursue self-discovery, including writing programs for experimentation and testing of conjectures. Solutions to many exercises are included at the end. The recommended course syllabus follows.

**** Syllabus**

- Introduction to scientific information and multidimensional visualization, complete course overview. One lecture with demos on the foundations and numerous applications.
- Projective geometry: foundations, dualities, collinearity theorems, homogeneous coordinates, linearization of projective transformations, conics in \mathbb{P}^2 , applications to computer graphics: transformations and their matrices; perspective: points at infinity are images of *ideal points*, geometrical invariants. Two lectures with demos.
- Parallel coordinates in the plane: dualities: *point* \leftrightarrow *line*, *translation* \leftrightarrow *rotation*, other transformations, parallelism and orthogonality, visual and automatic data mining. One lecture with demos.
- Lines in N -space; two lectures
 - Representing lines by $(N - 1)$ points with two indices, parallel lines, general representation, the 3-point collinearity property; demos.

- Transformations, improving the display of multidimensional lines, intersections, minimum L_1 distance between pairs of lines, upper and lower proximity bounds.
- Application: collision-avoidance algorithms for air traffic control; demos.
- Planes, hyperplanes, and flats in N -Space; three lectures.
 - Detecting coplanarity of points on a grid, representing hyperplanes by $(N - 1)$ parallel lines and a point (polygonal line), duality: *rotation of a plane about a line* \leftrightarrow *translation of a point on a line*, manufacturing data; demos.
 - Recognizing M -dimensional objects from their $(M - 1)$ -dimensional subsets, *recursive construction mapping*.
 - Detecting coplanarity of randomly chosen points, representing hyperplanes by $(N - 1)$ points with N indices, the family of *superplanes* and their special properties, reading the equation of a hyperplane from the picture, duality: *rotation of a plane about a line* \leftrightarrow *translation of $(N + 1)$ points on lines*, designing rotations and transformations from their “planar signature,” synthetic constructions: parallel planes, intersections, above/below relation between points and planes and more, “near coplanarity,” separating point-clusters on several planes; demos.
- Envelopes of one-parameter families of curves, background for the curves and surfaces chapters. One-half lecture (discretionary).
- Curves; one and one-half lectures.
 - Point-curves and line-curves, convex down/up curves and their transforms, duality: *cusp* \leftrightarrow *inflectio point*, *point-curve* \leftrightarrow *line-curve*.
 - *conics* \leftrightarrow *conics* via the Möbius transformation; six cases, algebraic curves: the Plücker formulas and dualities, curve plotters, examples of curve transforms, analysis of important points and portions of curves, periodicities, and symmetries.
 - Convex curves and their relatives, generalized conics and their dualities; six cases, operational transformations.
- Families of proximate hyperplanes; one and one-half lectures.
 - Topology for proximity of flats.
 - Proximity of lines and line neighborhoods.
 - Proximity of hyperplanes, formulation of the problem in N -space, hyperplane neighborhoods with $(N - 1)$ regions, the three cases: bounded

- convex polygons, unbounded convex polygons, generalized hyperbolas, constructing neighborhoods with complexity $O(N)$, the general case.
- Navigation within neighborhoods, finding the $(N - 1)$ points representing specific hyperplanes, matching algorithm; demos.
- Applications, a central problem in many fields, statistics (regression), geometric modeling, computer vision.
- Hypersurfaces in N -space; two lectures.
 - Formulation of the problem, hypersurfaces represented by $(N - 1)$ *linked* regions, boundary contours, quadric and general algebraic hypersurfaces.
 - Developable hypersurfaces represented by $(N - 1)$ linked curves, ambiguities and uniqueness, reconstruction, classes of developables, duality: *space curves* \leftrightarrow *developables*, examples: developable helicoids and more; demos.
 - Ruled hypersurfaces, several examples of their representation including Möbius strip, visualizing nonorientability in N dimensions; demos.
 - More general hypersurfaces, convex and nonconvex, and how to distinguish them in N dimensions, searching for matching algorithms, interior points.
 - An important approximate representation, interior points, applications to decision support and intelligent process control; demos.
- Data mining and other applications (self-contained chapter); two lectures.
 - Origins, the case for visualization.
 - exploratory data analysis (EDA); visual data mining
 - * Multidimensional detective, atomic queries, an easy case study: GIS data, Boolean operators and compound queries; financial data, hundreds of variables; demos.
 - Automatic classification, feature selection and ordering, comparisons, and two case studies; demos.
 - Visual and computational models, decision support: interrelationships, sensitivities, constraint and trade-off analysis; demos.
- Recent developments (appendix); two lectures (discretionary).
 - Representing several lines efficiently; Shlomi Cohen-Ganor.
 - Separating points on several planes; Nir Shahaf.
 - Surface representation decomposition and developable surfaces; Chao-Kuei Hung.

- Network visualization and analysis with $\|\cdot\|$ -coords; Yaron Singer and Ohad Greenspan.
- To see \mathbb{C}^2 : visualizing complex-valued functions; Yoav Yaari.
- Time permitting, students can choose to lecture on various topics in visualization, e.g., Internet visualization, fluid flow, visualizing the function of neural networks, detecting network intrusions, etc.

Prerequisites

The course attracted students from mathematics, computer science, engineering, statistics, management, data mining, and a wide variety of sciences (physics, chemistry, biology, etc.), geography, social sciences, and even linguistics. They ranged from second year to advanced doctorate students. Background in linear algebra is important for those who want to master the subject. Programming skills including some computer graphics are helpful though not necessary.

Acknowledgments

I first proposed parallel coordinates in 1959 when I was a graduate student at the University of Illinois (Champaign-Urbana). Two topologists, Professors D. Bourgin and S.S. Cairns, encouraged me to pursue the idea, but it took me many years to take their advice seriously. Later, I had the pleasure and privilege to collaborate with Bernie Dimsdale⁴ at the IBM Los Angeles Science Center (LASC), who made many important contributions and left his imprint throughout this work. Also at LASC, Alex Hurwitz's numerous contributions, suggestions, and ideas as well as Juan Rivero's application to the phase space for nonlinear differential equations benefited the early development.

From the students, starting in 1978 Moti Reif was bravely the first to do his graduate research on parallel coordinates (convexity algorithms) [110]; John Eickemeyer discovered among other things the "superplanes" [47]; Avijit Chatterjee developed Eickemeyer's results further for the visualization of polytopes in N -space [23]; Misha Boz made many contributions over the years, and together with Bernie Dimsdale, we did the first successful implementation of the collision avoidance algorithms [104]; Steve Cohan [31] and Paolo Fiorini developed applications to robotics [60]; Jim Adams (circa 1981), Alex Hurwitz, and later Tuval Chomut pioneered the first successful implementation of parallel coordinates for exploratory data analysis (EDA) [30] as well as contributing to the convexity algorithms with Moti Reif; Chao Kuei Hung derived an elegant characterization of developable surfaces in terms of their tangent planes [93], [108] (see [94] for updated results); Lou Ann Colby investigated dualities between rigid transformations; Richard Addison later improved the implementation of the collision avoidance algorithms; and Greg Wong together with John Eickemeyer under Alex Hurwitz's leadership developed and implemented further substantial improvements; Tova and Shlomo Avidan made innovative improvements on the data mining application [5], [103]; Tanya Matskewich and Yaakov Brenner did a very sophisticated and useful characterization of proximity for flats [131] that developed into a full chapter; Nadav Helfman [81] further developed the application to decision support; Adi Shlomi designed and wrote much of the software in the *ILM*. Liat Cohen solved the exercises and also contributed many ideas and improvements. David Adjashvili wrote the

⁴A long-term associate of John von Neuman.

beautiful software displaying surfaces in \parallel -coords. In the spirit of geometry, these pictures led to key results that were subsequently proved. Several other student projects made important innovations and are acknowledged in the text.

Many researchers contributed to the development and applications of parallel coordinates (in alphabetical order). The Andrienkos [2], J. Dykes et al. [44], R. Edsall [45], and A. MacEachren et al. [43] introduced \parallel -coords to GIS and geovisualization, Kim Esbensen [53] contributed to data analysis and dualities, H. Carter, C. Gennings, and K. Dawson on response surfaces in statistics [67]; Amit Goel [71] on aircraft design; Chris Jones in optimization [115]; John Helly [82] on early application to data analysis; Hans Hinterberger's [156] contributions in comparative visualization and data density analysis; Matt Ward et al. introduced hierarchical parallel coordinates [65] and added much to the field [181], Helwig Hauser's innovative contributions included parallel sets for categorical data [79]; Antony Unwin et al. made wide-ranging numerous contributions [177]; Li Yang [187] applied \parallel -coords to the visualization of association rules; H. Choi and Heejo Lee's [29] and P. Hertzog [83] are important contributions to detecting network intrusion, other important contributors to the field include Luc Girardin and Dominique Brodbeck [70], Xiabin Yen, et al. [189], T. Itoh, et al. [111], and Stefan Axelsson [6]. G. Conti [32] produced a ground-breaking book on security data visualization; there are exciting recent works by H. Ye and Z. Lin's [188] with an innovative contribution to optimization (simulated annealing); T. Kipouros et al. [120] with a sophisticated optimization for turbomachinery design [120]; S. El Medjani et al. [49] proposed a novel application using \parallel -coords as straight-line detectors; R. Rosenbaum and H. Schumann [151] on progression in visualization; F. Rossi [153] on visual data mining and machine learning; Huamin Qu et al. [89] on air pollution analysis and clustering; M. Tory et al. on \parallel -coords interfaces [171]; J. Johanson et al. [113]; G. Ellis and A. Dix [50] on clustering and clutter reduction; H. Siirtola and K.J. Räihä on interaction with \parallel -coords; B. Pham, Y. Cai and R. Brown on traditional Chinese medicine [147], and there are proposals to enhance \parallel -coords using curves [129], place them in 3-D [139], [114], or modify them as starplots [51]. C.B. Hurley and R.W. Olford contributed to the definitive study on axes permutations [95]. This list is by no means exhaustive.

There have been a number of noteworthy visualization contributions since the 1980s: the grand tour by Asimov [4], dimensional inference and more by Buja et al. [66], multidimensional scaling and convergence of correlation matrices by Chu-Hou Chen [27], [26], a chain of valuable and varied ideas (most notably RadViz) by Georges Grinstein et al. [87] [88] and Daniel Keim et al. [119] [118], hierarchical visualization of multivariate functions by Mihalisin et al. [135], [84]. Eminent are the seminal contributions of Jacob Zaavi and Gregory Piatetsky-Shapiro et al. [57] in all aspects of data mining and the role of visualization.

Bertin's book [14] is the early classic in *information visualization*. The field has been blossoming since then as attested by the volumes [22] and [11] containing organized collections of important papers, the seminal source on data mining [57], and the wonderful texts of Colin Ware emphasizing perception and design [182], Chaomei Chen's [25] work on a range of cognitive social and collaborative activities, and the comprehensive expository collection of methodologies by Robert Spence [164].

I have been supported and introduced to a variety of good problems by several friends and colleagues, especially (listed in alphabetical order) by George Bekey, Ronald Coiffman, Julia Chung, Alon Efrat, Marti Hearst, Peter Jones, Pei Ling Lai [109], [124], Khosrow Motamedi, Karl Müller, David McIlroy, Niko Shlamberger and Ana Trejak, Randall Smith of GM Research, and Fred Warner. I am especially grateful to John Nacerino of EPA in Las Vegas for his guidance and steadfast encouragement. His initiative and efforts enabled the distribution of the parallel coordinates software to academic researchers. Since 1990 the IEEE annual visualization conferences have been an excellent forum in which to learn and present ideas as well as receive valuable suggestions and criticism. In this community the following colleagues were especially helpful: Dan Bergeron [185], Ed Chi [28], Georges Grinstein, Beth Hetzler, Helwig Hauser, Daniel Keim, Chaim Lefkovits, Ted Mihalisin, Tamara Munzer, Alex Pang, Catherine Plaisant [160], Bill Ribarsky, Bernice Rogowitz, Holly Rushmeier, Lloyd Treinish, Bob Spence and Lisa Tweedie [175], John Stasko, Theresa-Marie Thyne, Matt Ward, and Pak C. Wong, who together with R.D. Bergeron wrote the first survey of the field [185] among their other contributions. I learned a great deal from Ben Shneiderman's [159] [160], [11] fundamental insights and encyclopedic contributions to visualization, and also from Marti Hearst's [112], [80] courses, varied interests and stimulating ideas in information visualization, user interfaces, and text mining. Of great benefit has been the interaction with the statistics and statistical graphics communities, especially with Daniel Asimov, Andreas Buja, Jogesh Babu, Dianne Cook [34], Larry Cox, Daniela Di Benedetto [33], Michael Friendly [62], Heike Hoffman, Catherine Hurley, Moon Yul Huh [91], Sigbert Klinke, Carlo Lauro [125], Jungi Nakano [139], Francisco Palumbo et al. [125], David Scott, Valentino Tontodonato [170], Ritei Shibata et al. [123], Deborah F. Swayne [168], Martin Theus, Luke Tierney, Simon Urbanek, Antony Unwin [176], Bill Venables, Ed Wegman [184], Adalbert Wilhelm, Lee Wilkinson and Graham Wills.

Special thanks for their help and encouragement are due to several of my colleagues in Israel: Yoav Benjamini, Iddo Carmon, Daniel Cohen-Or, Ronen Feldman, Camil Fuchs, Dony Gat, Hillel Gauchman, David Gilat, David Levin, Mark Last, Leonid Polterovich, Eugenii Shustin and Jacob Zahavi.

Over the years, many managers at IBM, among them Lew Leeburg, Baxter Armstrong, Homer Givin, Jim Jordan, John Kepler, Kevin McAuliffe, Abe Peled, Bernie Rudin, Arvid Shmaltz, and John Wolff, supported, encouraged and in many ways helped with this effort.

With patience Ann Kostant, Laura Held, Alicia de los Reyes, and their staff from Springer most capably encouraged, cajoled, corrected and produced this book. Special thanks are due Chao-Kuei Hung, who read the whole manuscript and provided invaluable ideas and corrections. To all these friends and others whom I may have inadvertently omitted, I offer my profound gratitude. I am especially grateful to my family for their unswerving *multidimensional* encouragement, inspiration, patience, and precious sense of humor; without them I would have never finished.

Introduction

1.1 Multidimensional Visualization

About half of our sensory neurons are dedicated to vision,⁵ endowing us with a remarkable pattern-recognition ability. Visualization, an emerging field with huge potential, aims to incorporate this tremendous capability into our problem-solving process. Rather than dictionary definitions we consider visualization as a collection of application-specific mappings:

$$Problem\ Domain \longrightarrow Visual\ Range.$$

For various classes of problems, *visual models* are constructed that provide *insight through images* [132] about the problem. This is an extension of Hamming's 1962 adage that "the purpose of computing is insight, not numbers." While the development of software and hardware tools has been impressive, concepts broadening the applicability of visualization have lagged behind. *Scientific visualization* for applications having a physical or geometric basis (e.g., electromagnetic fields, geography, fluid flow) has blossomed. Yet progress has been slower in *information visualization* for fields not endowed with such structure, as in finance, process control, and communications, where many variables are involved, further complicating their visual expression.

⁵If one takes as an informal but meaningful definition of visual as "affected by light stimulation to the eyes," and if the quantity of neurons in the cortex is used as a baseline, then the 50% figure is the one often quoted by experts. I am grateful to Prof. Enio Mingolla, of Boston University, and Prof. Simo Vanni, of Helsinki University of Technology, who together with other experts confirmed this information.

Understanding the underlying geometry of multidimensional problems can provide insights into what is possible and what is not. In an astonishing but not well known paper [46] the physicist Ehrenfest showed in 1917 that planetary orbits are stable only in space of dimension 3. Higher-dimensional planetary systems, if they ever existed, would have a short career due to the orbits' instability, which offers an interesting hypothesis for the dimensionality of our habitat. Another dimensionality curiosity is that rotating rigid bodies have an axis of rotation only in odd-integer dimensions [169, p. 189], Goldstein [72, pp. 158–166]. Had our dimensionality been 4, the Earth would not have a fixed axis, resulting in the absence of seasons,⁶ among other things. Clearly dimensionality can have important implications. For a collection of anecdotes and opinions of great scientists on the role of visualization in science see Ron Kriz's website [122]. For “dimensional adventures,” the classic *Flatland* [1] by Abbot, *Flatterland* by Stewart [165] and Banchoff's journey beyond the third dimension [8] are warmly recommended.

The fascination and exploration of dimensionality may predate Aristotle and Ptolemy, who argued that space has only three dimensions. It was not until the nineteenth century that Riemann, Lobachevsky, Bolyai, and Gauss unshackled our imagination, limited by the experience and perception⁷ of our three-dimensional habitation, with their non-Euclidean and higher-dimensional geometries. Would a way be found to visualize these wonders? People have certainly tried, yet the available multidimensional visualizations can display only finite multivariate point sets (see the beautiful collection in [63] [172], [173], [174] and bibliographies in [185] and [106]).

1.2 *FT-I* How?

For the visualization of multivariate *relations* a conceptual breakthrough emerged that is, in the spirit of Descartes, a multidimensional coordinate system. In geometry, *parallelism*, which does not require a notion of angle, rather than orthogonality, is the more fundamental concept. Coupled with the fact that orthogonality “uses up” the plane very fast, this inspired me to experiment and then present⁸ *Parallel Coordinates* in 1959. The systematic development began in 1977 [97] (short reviews

⁶Resulting from the axis' tilt.

⁷As an interesting aside, there is strong evidence suggesting that even 3-D perception depends greatly on our 2-D cognition (see [20], for example).

⁸In his combinatorial topology course, Prof. S.S. Cairns lamented the absence of multidimensional visualization tools, prompting me to share the idea with the class.

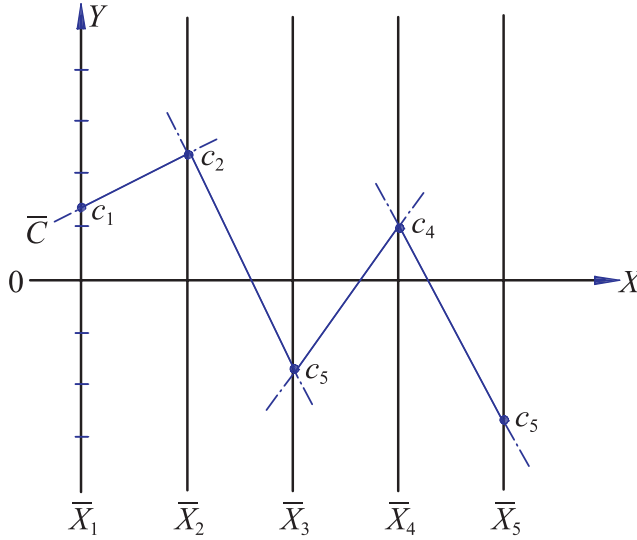


Figure 1.1. The polygonal line \bar{C} represents the point $C = (c_1, c_2, c_3, c_4, c_5)$.

in [105], [100]). The goal was the systematic visualization of multidimensional problems without loss of information having the properties listed below.

1. It should have low representational complexity, i.e., $O(N)$ for N variables, be linear in the number of dimensions (variables); compare with the common “scatterplot matrix,” whose pairwise projections have $O(N^2)$ complexity.
2. It should apply to any dimension, treating every variable uniformly, unlike *Chernoff faces* or *glyphs* and other multivariate visualizations [172].
3. The displayed objects should be recognizable under projective transformations (i.e., rotation, translation, scaling, perspective).
4. Properties of the N -dimensional objects should be apparent from the display.
5. It should be mathematically rigorous.

These points are amplified in Chapter 10 on data mining.

In the Euclidean plane \mathbb{R}^2 with xy Cartesian coordinates, N copies of the real line \mathbb{R} labeled $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N$ are placed equidistant and perpendicular to the x -axis. They are the axes of the *parallel coordinates* system for the Euclidean N -dimensional space \mathbb{R}^N , all having the same positive orientation as the y -axis. A point $C \in \mathbb{R}^N$ with coordinates (c_1, c_2, \dots, c_N) is represented by the *complete* polygonal line \bar{C} (i.e., the complete lines containing the segments between the axes) whose N vertices are at $(i - 1, c_i)$ on the \bar{X}_i -axis for $i = 1, \dots, N$, as shown in

Fig. 1.1. In this way, a one-to-one correspondence between points in \mathbb{R}^N and planar polygonal lines with vertices on the parallel axes is established.

In principle, a large number of axes can be placed and be *seen* parallel to each other. The representation of points is deceptively simple, and much development with additional ideas is needed to enable the visualization of *multivariate relations* or equivalently multidimensional objects. Specifically, the representation of an M -dimensional object, $2 \leq M$, will be obtained from its $(M - 1)$ -dimensional subsets. For example, the representation (i.e., image) of a line is obtained from the polygonal lines representing *points* on the line, and this turns out to be their envelope.⁹ Next, the representation of a plane in \mathbb{R}^3 is obtained from the representation of the *lines*, rather than the points, it contains. This leads to a recursion that turns out to work splendidly, but we are getting ahead of ourselves. The *recursive construction* algorithm is properly considered in Chapter 5 on hyperplanes.

Parallel coordinates (abbreviated \parallel -coords) transforms the search for multivariate relations in a data set into a pattern-recognition problem. This is the core reason for the widespread use in exploratory data analysis (EDA) (also called “visual data mining”) and the inclusion in software packages starting with EDA (Chomut [30]), [59], (Hinterberger [156]), Spotfire (Christopher Ahlberg and Ben Shneiderman [3]), ExplorN (Bolorfoush, Carr et al.), WinViZ (Eickemeyer), VisDB ([119]), Xmdv ([181]), XGobi ([167]), Strata ([155]), PVE (Adams, Hurwitz, Inselberg, Chatterjee [9]), and many others. This type of application [101] hinges on:

- an informative display of the data,
- good choice of queries, and
- skillful *interaction* between the user and display in searching for patterns corresponding to relationships among the variables.

Part of the discovery process (classification) has been automated [103], and the *Parallax* software contains both visual and automatic data mining modules based on \parallel -coords. The self-contained Chapter 10 is devoted to the *data mining* applications, the design of queries, guidelines, strategies, and “tricks” for knowledge discovery and classification. They are illustrated with several case studies on real multivariate (one with hundreds of parameters) data sets garnering surprising and at times “spicy” information. There are other applications to statistics [67], air traffic control (collision avoidance algorithms, presented in Chapter 4 on lines), geometric modeling and computer vision (Chapter 8 is on the visualization of families of “close” lines and hyperplanes), approximations, decision support (including process control) [81], optimization ([24], [115]), recent different innovative applications like [178]

⁹Chapter 6 contains a short intuitive introduction to the theory of envelopes.

and others acknowledged earlier. For *data mining*, *the grand vision* is the creation of algorithms whose input is a massive data set and output 2-dimensional patterns; the “graph” of multivariate relations discovered in the data and, as described in Chapter 10, this goal is now within our grasp.

The next chapter provides background in geometry. Readers may skim through it, referring to specific topics as the need arises. Prior to that, some exploration on the Internet is worth doing with queries such as:

- Visualization
- Multivariate + Visualization
- Multidimensional + Visualization
- “Scientific Visualization”
- “Information Visualization” or Infovis
- “Parallel Coordinates”
- “Parallel Coordinates + Applications”
- Visualization + Applications
- “Parallel Coordinates” + a specific application.

This page intentionally left blank

2

Geometry Background

“Let no one ignorant of Geometry enter” — At entrance to Plato’s Academy

2.1 Preliminaries

2.1.1 In the Beginning ...

The Nile’s flooding in Egypt used to destroy landmarks and position markers, raising the need for reliable measurements and originating the development of geometry *Geo-Γεω* (for land) *metry* (for measure-μετρία).¹⁰ The Rhind Papyrus, which is in the British Museum, was written by an Egyptian priest named Ahmes about 1,800 B.C.E. It treats the measurement of planar regions and solids. It contains the remarkable approximation of $\pi \approx \left(\frac{16}{9}\right)^2 = 3.1604$. The contemporary Babylonians, as well as Chinese, used 3 for the value of π . The value of 3 for π is also mentioned in the Hebrew scriptures (see I Kings 7:23). A rabbi called Nehemia (p. 5 [74]) proposed the value of $\frac{22}{7}$ for π , but apparently it was not accepted. By the way, only in 1789 was it first proved by J. Lambert that π is an irrational number and in 1882, F. Lindemann proved that π is transcendental (i.e., it is not the root of any algebraic equation with rational coefficients). For more enjoyable folklore on π the reader is referred to [12].

It is fair to consider the Egyptians more as surveyors, who used a collection of geometrical formulas, mostly approximations, rather than geometers. For proper attribution, *deductive* geometry was accomplished by the Greeks, probably beginning with Thales of Miletus (\approx 640–545 B.C.E.), who was familiar with Egyptian

¹⁰This story was reported by Proclus in the fifth century C.E., who wrote the early history of geometry.

and Babylonian mathematics. Pythagoras, Hippocrates (not the physician of the same name), Menaechmus, Archimedes, Apollonius, Plato, Archytas, Eudoxus and Theaetetus continued the development until it was codified by Euclid in his *Elements* $\Sigma\tau\omicron\iota\chi\epsilon\iota\alpha$ (≈ 300 B.C.E.). Entirely novel, and characteristic of Greek mathematics, is the *axiomatic method* whereby theorems are established by proof. Particularly relevant to computer science are the *constructions* of Euclidean geometry. They consist of a finite set of well-defined operations and provide an unambiguous procedure that terminates for all inputs! Truly the first *algorithms ... and with proofs*. The constructions of Euclidean geometry involved ruler and compass. A remarkable discovery was made in 1672 by the Danish geometer G. Mohr (1640–1697) and independently in 1797 by the Italian L. Mascheroni (1750–1800).¹¹ They showed that *all* of these constructions can be accomplished by *compass alone*! [38]. The proofs were, in the true spirit of computer science, by simulation. It was demonstrated that any operation involving a ruler can be replaced by a finite number of compass operations; a real equivalence proof like those in automata theory. Alternatively, J. Steiner and J.V. Poncelet showed that all constructions in Euclidean geometry can be accomplished with a straightedge when in addition, a single circle and its center are also given. Continuing in this vein, it is natural to ask what is left if the compass is discarded and only the ruler is left. What remains is *projective geometry*, an elegant branch of geometry simpler than Euclid's but rich and interesting in its own right. It deals with points, lines, planes, and spaces where distance between two points or angles between two lines are *not* measured and there are *no parallel lines* in the plane! For a modern treatment of projective geometry see [154].

Exercises

1. To get a feel for the Mohr–Mascheroni result, using compass *only* write (and prove) an algorithm for:
 - (a) constructing 3 collinear points,
 - (b) given a segment \overline{OA} construct a point B such that \overline{OB} is twice as long as \overline{OA} ,
 - (c) bisecting a segment,
 - (d) trisecting a segment,
 - (e) given a circle C and two points determining a line l , construct the point of intersection $C \cap l$ (be careful here, the line l is *not given*, but only two points on the line l are. This is what makes the problem *hard*).
 - (f) constructing the point of intersection between two lines (*hard*).

¹¹For a short exposition on the *geometry of compasses* see [145, (pp. 23–25)] and [36, (pp. 145–151)], quoted in [38].

2. Given an angle A , provide a construction for finding an angle $= 1/2 A$ using the Steiner–Poncelet constraint, i.e., with a straightedge, a single circle, and its center.

To recall and rejuvenate some Euclidean geometry skills try some of the following¹²:

3. Provide a Euclidean geometry construction (i.e., using circle and ruler only) that transforms a rectangle into a square of equal area.
4. One of the three famous unsolvable problems in Euclidean geometry is “squaring the circle.” Given a square with area A construct, using ruler and compass only, a square having area $2A$. What other shapes can this be generalized to?
5. Prove that a triangle ABC is isosceles if the bisectors AD and BE of the two angles A and B are equal. Here D and E are the points of intersection of the angle bisectors with the sides BC and AC respectively. This is a deceptively *hard* problem.

2.2 ♣ **FT-1** Why Projective Geometry?

We begin the formal aspects with an excursion into projective geometry, but why? There are two important concepts and tools that we will need and that are also widely used in many other fields and applications. They are:

- *duality* and
- *homogeneous coordinates* (see [146] for applications of homogeneous coordinates to computer graphics).

Both of these are projective concepts and can be properly understood only in this context. The experience is quite broadening and will get us “unglued” from the narrow mindset of Euclidean geometry. For example, entirely new notions, such as finite geometries, arise naturally. Projective geometry was initially motivated by the fine arts. The Italian architect Brunelleschi in 1425 first proposed a geometrical theory of perspective that was formalized by Alberti a few years later. It included the idea of “points at infinity” and influenced art in a profound way, as shown by contrasting prior “flat” paintings such as Martini’s *Annunciation* with those where “depth” is perceived, such as Leonardo’s *Last Supper* and Raphael’s *Athens*, shown in the chapter on projective geometry by Morris Kline [140, vol. 1, 622–641]. In the absence of photography, realism in drawings had not only artistic but also great practical value.

¹²See also the delightful short paper by Hudson [90].

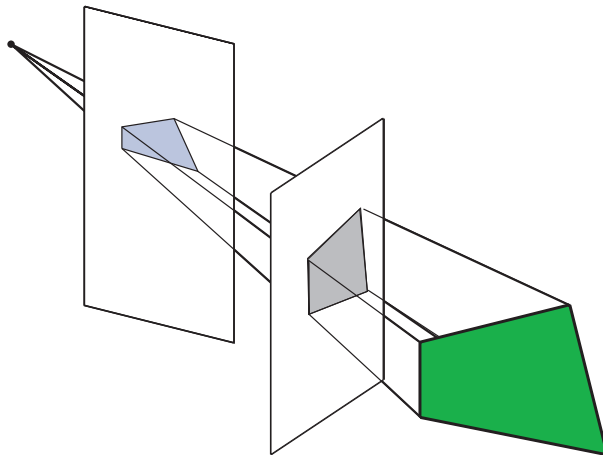


Figure 2.1. Projections and Sections.

2.2.1 Projective Plane Model

Projective geometry deals with properties that are invariant under projections and sections, see Fig. 2.1, such as “incidence” and “collinearity.” A beautiful projective theorem, discovered by the French architect Desargue in 1640, states that two triangles, ABC and $A'B'C'$, are in perspective from a point O if and only if they are in perspective from a line TSR ; see Fig. 2.4 (left) in the Euclidean plane and Fig. 2.4 (right), where O is a “point at infinity” in the vertical direction. Note the role of *duality* in the interchange between the words *point* and *line*. Of course, we have not explained what is meant by “perspective”; see Fig. 2.2 for an indication, but we are on the way.

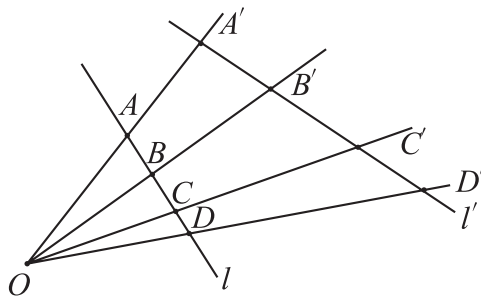


Figure 2.2. Points and lines in Perspective.

It will help our intuition to have a model for visualizing the projective space, and only afterward do we present the more formal axiomatic treatment. Our “building blocks,” otherwise called *primitives*, are concepts and objects with which we are already familiar. The notation commonly used in geometry appears in the list below; points and lines are denoted by capital and lowercase Latin letters (preferably in *italics*) respectively, while planes are denoted by lowercase Greek letters.

Primitives

- *points* - P, Q, A
- *lines* - ℓ, r, s
- *planes* - π, α, β
- *incidence relation* - P is *incident* to ℓ , or P is *on* ℓ .

Of course, incidence and containment are not equivalent, since P is *incident* to ℓ if and only if ℓ is *incident* to P ; $P \in \ell$ is true and $\ell \in P$ is not.

We start with the Euclidean space \mathbb{R}^3 and add something new. Let $\pi \subset \mathbb{R}^3$ be a plane. To each line $\ell \subset \pi$ we associate an *ideal point* (“point at infinity”) L_∞ such that distinct intersecting lines have distinct ideal points while parallel lines have the *same* ideal point. Effectively, then, ideal points signify direction. Since no

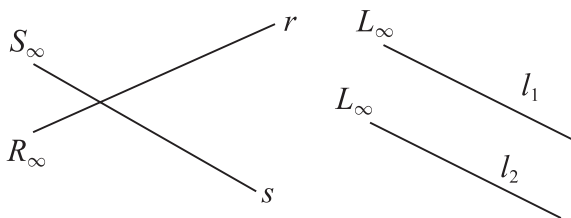


Figure 2.3. Ideal points and augmented lines.

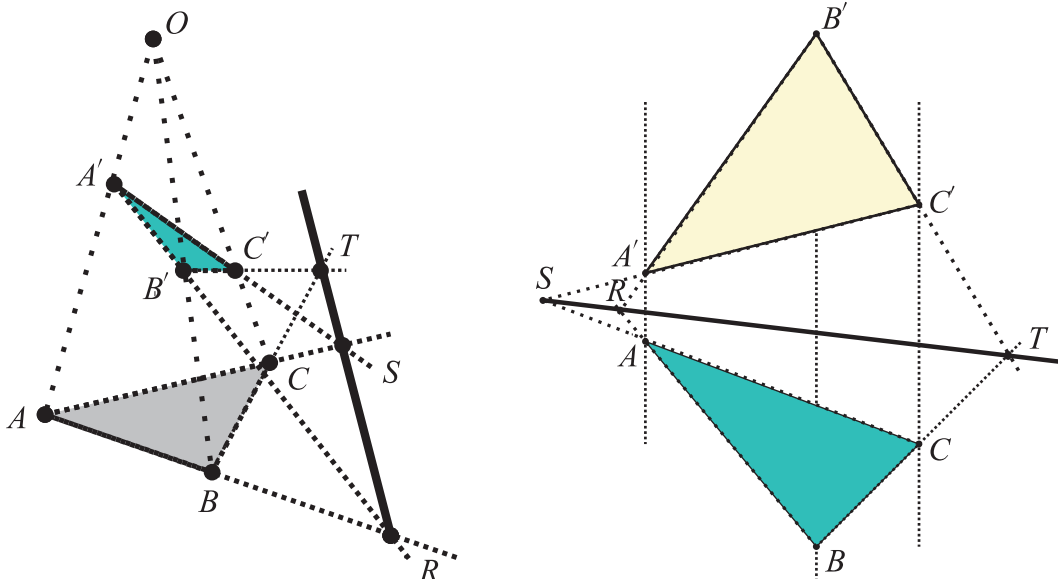


Figure 2.4. Desargue's theorem. The three points T, S, R , $T = B'C' \cap BC$, $S = A'C' \cap AC$, $R = A'B' \cap AB$, are always collinear for when the triangles are in perspective with respect to an ordinary point (left) or an ideal point (“point at infinity”) (right).

distinction is made between the “+” and “−” directions for each line ℓ , one and not two L_∞ need to be appended.

On occasion, the notation P_m^∞ is used for ideal points, with m being the slope of the corresponding direction. The points of \mathbb{R}^3 are called *regular* to distinguish them from the “*ideal*” ones. Proceeding, we call the pair (ℓ, L_∞) an *augmented line*. Then for $r, s \subset \pi$,

$$(r, R_\infty) \cap (s, S_\infty) = \begin{cases} P \text{ ordinary point} & \iff r \text{ not } \parallel s, \\ R_\infty & \iff r \parallel s. \end{cases}$$

For a given plane $\pi \subset \mathbb{R}^3$, let $p_\infty = \{L_\infty | \ell \subset \pi, (\ell, L_\infty)\}$ be the collection of all ideal points (directions) on π . We will see that p_∞ satisfies the axioms for a line, and for this reason it is called an *ideal line*. The pair (π, p_∞) is called an *augmented plane* for, as will be seen, it satisfies the axioms for a plane. Proceeding, the ideal lines for each plane $\pi \subset \mathbb{R}^3$ are collected and denoted by $\pi_\infty = \{p_\infty | \pi \subset \mathbb{R}^3, (\pi, p_\infty)\}$; we refer to π_∞ as the *ideal plane* and the pair $(\mathbb{R}^3, \pi_\infty)$ as the *augmented space*. All in all, then, “points” are either ordinary or ideal, “lines” are either augmented or ideal, and “planes” are either augmented planes or the ideal plane.

Hilbert [85] proposed an elegant model of the projective plane whose construction (actually a modified version) is shown in Fig. 2.6. A hemisphere S together with the disk D capping it is placed in a Euclidean plane π , with its south pole tangent to the origin. A *stereographic* projection is established such that a point $P \in \pi$ is connected to the center C of S . The point $P' = CP \cap S$ is the image of P . All points of π (i.e., “ordinary points”) are mapped uniquely onto the surface points of S . As the point $P = (a_1, a_2)$ goes farther and farther away from the origin in the constant direction having slope $m = a_2/a_1$, it approaches the ideal point¹³ P_m^∞ . Then the line CP approaches the line parallel to π that includes the diameter d_m of the disk D having slope m . This establishes a one-to-one correspondence $d_m \longleftrightarrow P_m^\infty$. The points on the hemisphere’s surface are the images of the ordinary points, while the diameters of D are the images of the ideal points providing the image of all of \mathbb{P}^2 as described above.¹⁴ It is a very helpful visual model. For example, a line $\ell \subset \pi$ is mapped into a great semicircle (see exercise below), and the augmented line $(\ell, \mathbb{P}_m^\infty)$ includes the diameter d_m . Parallel augmented lines, as shown in Fig. 2.7, “intersect” at their common ideal point. Two points on a projective line specify *two*

¹³The subscript m , in the notation of an ideal point, indicates the corresponding direction, and it is included only when the direction is relevant.

¹⁴We use homogeneous coordinates only with real values so there is no need to use \mathbb{RP}^2 as seen elsewhere, the \mathbb{R} distinguishing the *real* projective plane from others.

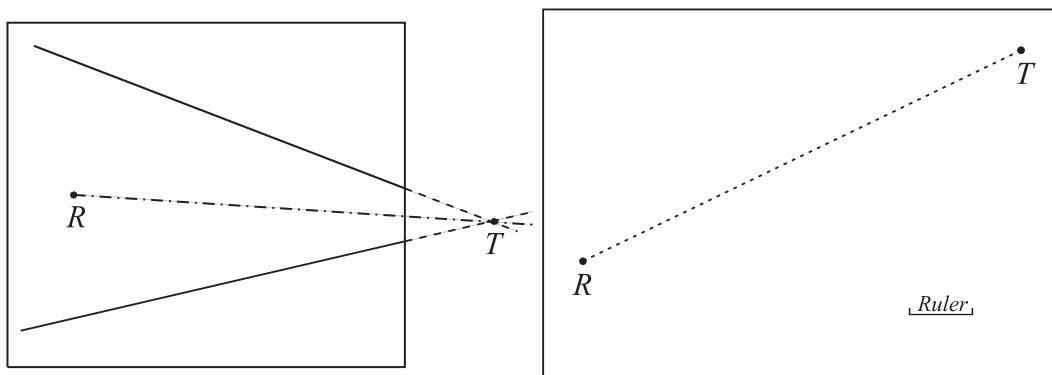


Figure 2.5. On the left, figure for Exercise 4; and on the right, figure for Exercise 5.

intervals; one containing the ideal point and its complement being the “ordinary” interval of the Euclidean line.

A very basic though often unfamiliar property of \mathbb{P}^2 is that it is not *orientable*. That is, it is not possible to assign a consistent sense of left and right ([18] p. 372). This matter arises in various topics in our development, and we shall learn how to cope with it when the need arises.

Exercises

1. Show that lines are mapped into great semicircles by stereographic projection.
2. Show that stereographic projection preserves angles (i.e., is isogonal)
3. How are circles mapped by stereographic projection?
4. On a piece of paper, as shown in Fig. 2.5 (left), there is a point R and portions of two lines intersecting at some point T outside the paper. Using only a straight-edge (ruler), draw the line joining the points R and T (Hint: Use Desargues’s theorem). From [13, vol. I, p. 119].
5. Given two points R and T and a short ruler, as shown in Fig. 2.5 (right), not long enough to reach both points, draw the line joining R and T (Hints: Use the result of the previous exercise and slide the ruler along a line segment.) From [13, vol. I, p. 119].

2.2.2 Axioms for Projective Geometry

The time has come to describe projective 3-space more formally in terms of the axioms, i.e., the properties of the primitives: points, lines, planes, and the *incidence* (or “*on*”) relation.

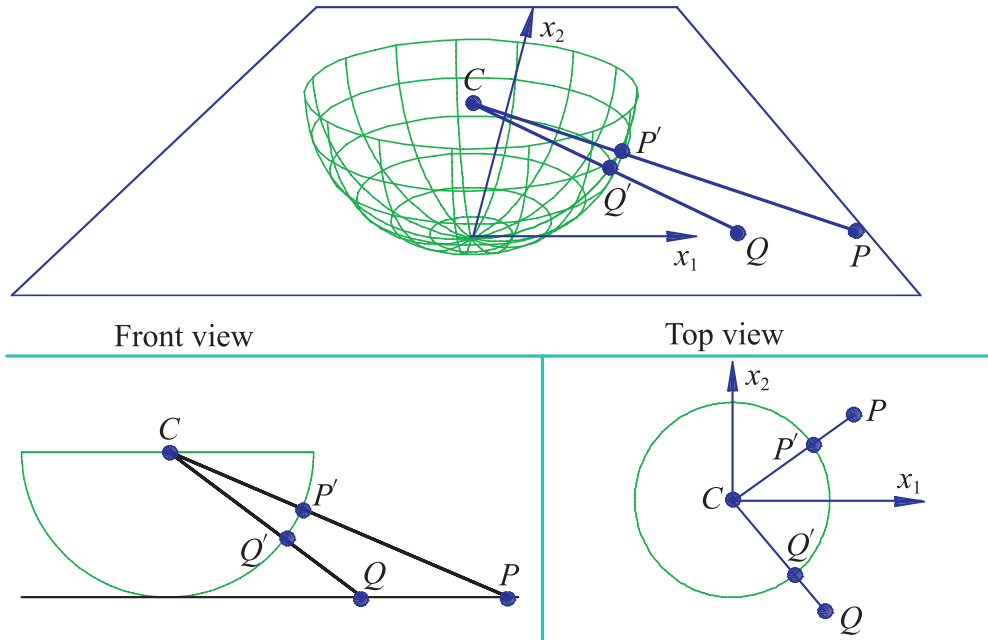


Figure 2.6. Model of the projective plane.

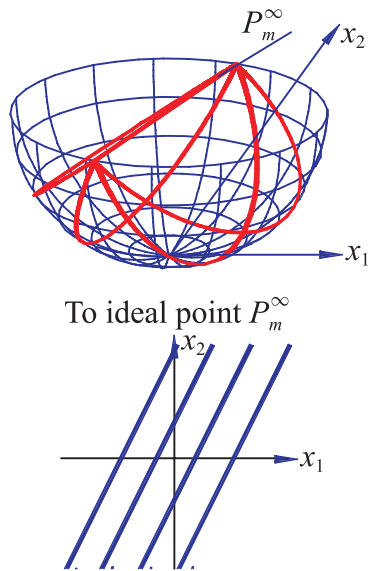


Figure 2.7. Parallel lines on the projective plane model.

Axioms for 2-D and 3-D

1. **(2-D)** Two distinct points are *on* (or incident to) a unique line. That is, if $P \neq Q$ then $P, Q \in \ell_1$ and $P, Q \in \ell_2 \Rightarrow \ell_1 = \ell_2$.

2. **(3-D)** Three distinct noncollinear points are *on* a unique plane. That is, if $P \neq Q, Q \neq R, P \neq R$, and there is no line $\ell \ni \{P, Q, R\} \subset \ell$, then $\{P, Q, R\} \subset \pi_i$, $i = 1, 2 \Rightarrow \pi_1 = \pi_2$.

3. **(* 2-D)** Two distinct coplanar lines are *on* (i.e., intersect or determine) a unique point. With the “ideal” elements we always have $\ell_1 \cap \ell_2 \neq \emptyset$ and $P_1, P_2 \in \ell_1 \cap \ell_2 \Rightarrow P_1 = P_2$.

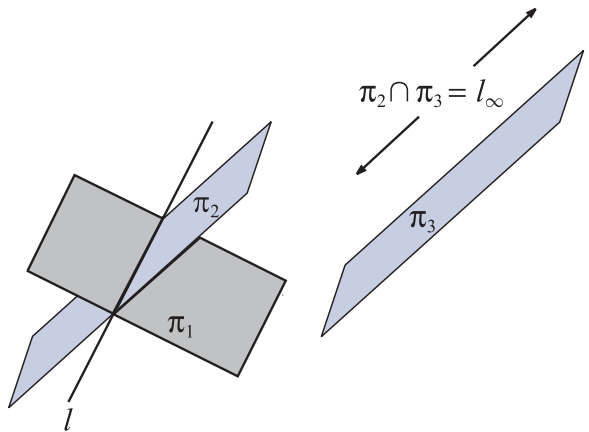
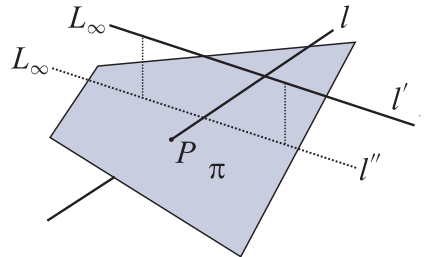
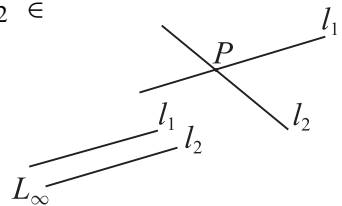
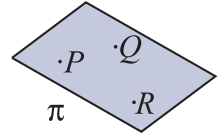
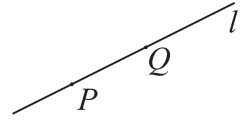
This axiom and the remaining axioms are marked with an asterisk (*), for they differ from their corresponding Euclidean cousins. Specifically, this axiom corresponds to Euclid’s famous fifth postulate, which postulated the existence of parallel lines.

4. **(* 3-D)** A plane π and a line ℓ not contained in π are *on* (i.e., intersect at) a unique point $\ell \cap \pi \neq \emptyset$ always and $P_1, P_2 \in \ell \cap \pi \Rightarrow P_1 = P_2$.

5. **(* 3-D)** Two distinct planes are *on* (i.e., intersect in) a unique line $\pi_1 \cap \pi_2 \neq \emptyset$ and $\pi_1 \cap \pi_2 = \ell$, where ℓ is unique.

It is not difficult to prove that the Euclidean plane together with the ideal and augmented elements satisfies the projective plane axioms. Let us assume that ideal points and augmented lines satisfy the axioms (see exercises) and study an example.

Theorem 2.2.1. *The ideal line p_∞ of an augmented plane (π, p_∞) is a line.*



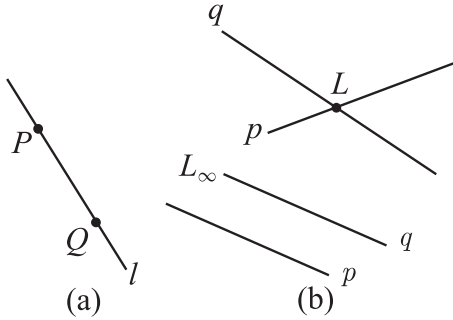


Figure 2.8. Axiom 1 is illustrated in (a) and its dual axiom (3) in (b).

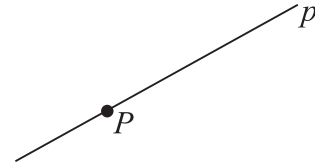


Figure 2.9. Self-dual: *point on line and line on point*.

Proof. We need to show that p_∞ satisfies the axioms for lines, namely axioms 3 and 4.

Let (ℓ, L_∞) be any augmented line on (π, p_∞) . Since p_∞ is the collection of all ideal points on the augmented plane, $(\ell, L_\infty) \cap p_\infty = L_\infty$, so axiom 3 is satisfied.

Let (π', p'_∞) be an augmented plane not containing p_∞ ; therefore π' is not \parallel to π . Hence $\pi \cap \pi' = \ell \Rightarrow (\pi, p_\infty) \cap (\pi', p'_\infty) = (\ell, L_\infty)$, where L_∞ is the ideal point associated with ℓ . ■

Two excellent references for projective geometry are [7] and [39].

Exercises

1. Show that the ideal points satisfy the axioms of points.
2. Show that augmented lines satisfy the axioms of lines.
3. Show that augmented planes satisfy the axioms of planes.
4. Show that the ideal plane π_∞ satisfies the axioms of a plane.

2.2.3 Principle of Duality

The axioms yield a principle not found in Euclidean geometry. In 2-D there is a point \leftrightarrow line duality stemming from (axiom 1) two *points* are on a *line* ℓ and (axiom 3) two *lines* are on a *point* L . As a result, planar figures can be “dualized” by switching the words point and line. For example, three distinct noncollinear (i.e., not on the same *line*) *points* P , Q , and R determine a triangle. Dually, three distinct nonintersecting (i.e., not on the same *point*) *lines* p , q , r also determine a triangle. That is, triangles are self-dual. Note also the projective cases in Fig. 2.11: (left)

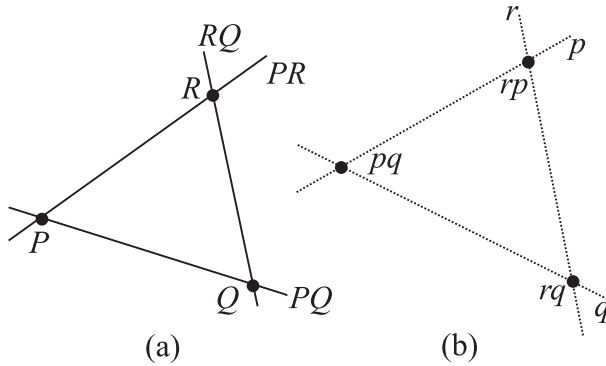


Figure 2.10. Self-dual 2: Triangle (a) and “triline” in (b).

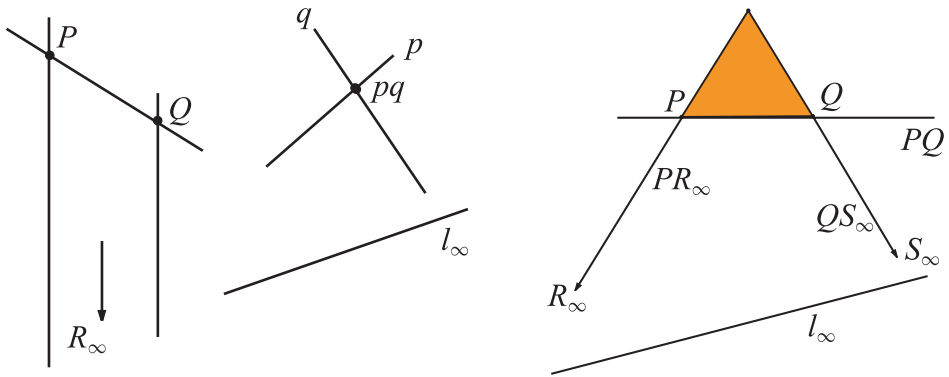


Figure 2.11. Projective triangles (left) and quadrangle (right).

triangles and (right) the four points P , Q , R_∞ , S_∞ , which form a quadrilateral (“quadrangle”) with edges PQ , PR_∞ , QS_∞ , l_∞ .

The 3-D duality is point \leftrightarrow plane and results from axiom 2, whose dual is that three noncollinear *planes* π_1, π_2, π_3 are on a point P ; Fig. 2.12. By the way, this is the reason why point \leftrightarrow line dualities do not generalize to higher dimensions. Other examples of the point \leftrightarrow plane duality are that two *planes* π_1, π_2 are on a line, whose dual is axiom 1, and two *points* P_1, P_2 are on a line. Also axiom 4: a *plane* π and a line l not on π are on (determine) a *point* is dualized to a *point* P and a line l not on P are on (determine) a *plane*.

The implications of duality are far deeper than those just indicated. For if a (projective) theorem involves points and lines in 2-D or points and planes in 3-D, its dual is also true. We illustrate this with the earliest known projective theorem,

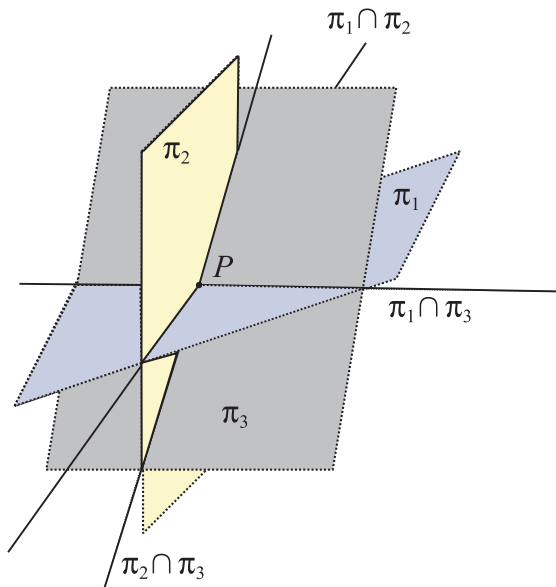


Figure 2.12. Three noncollinear *planes on a point*; dual of axiom 2.

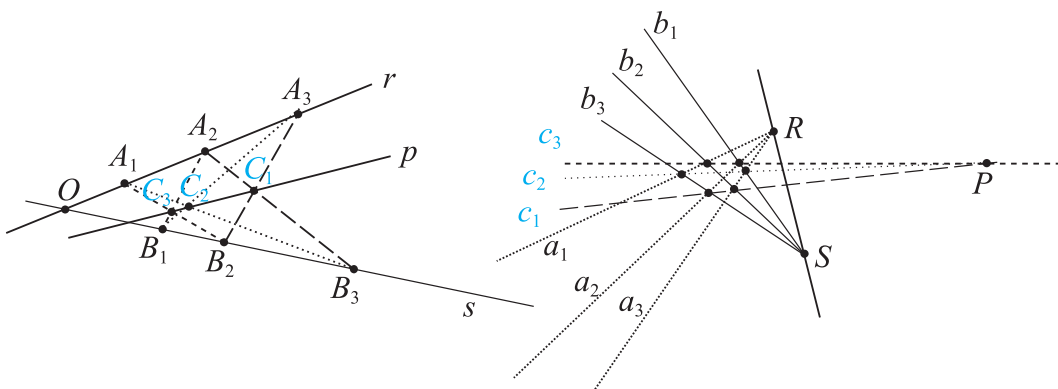


Figure 2.13. Pappus's theorem (left) and its dual (right).

proved by Pappus in the third century C.E. Using the notation \cdot , instead of \cap , for intersection Pappus's theorem states that for any distinct *points* A_i on a *line* r and any distinct *points* B_i on a *line* s , where $i = 1, 2, 3$, the *points* C_i , where $C_i = A_j B_k \cdot A_k B_j$ for $j \neq k \neq i$, are always on a *line* p . The dual theorem is obtained simply by interchanging the *words* *point* \leftrightarrow *line*. In our notation it is

equivalent to switching lowercase \leftrightarrow capital letters as follows: for any distinct *lines* a_i on a *point* R and any distinct *lines* b_i on a *point* S , where $i = 1, 2, 3$, the *lines* c_i , where $c_i = a_j b_k \cdot a_k b_j$ for $j \neq k \neq i$, are always on a *point* P . No further proof is needed!

Fun and Games with the ILM

Open the ILM: set *Chapter* = 2D Point Line duality.
 The display shows two sets of axes.
 On the right, click to draw a point and see the corresponding line on the left
 Move the point and watch the line.
 Next set *Subject* = Point \leftrightarrow Line.
 Draw a point on the left and see the corresponding line on the right.

2.3 ** Finite Geometries

Let a *finit* set of points and a nonempty collection of point subsets called *lines* satisfy the following axioms:

1. Any two distinct points A and B are *on* a unique line ℓ .
2. If p and q are distinct lines, there is a unique point *on* both.
3. There are at least three distinct points *on* any line.
4. Not all points are *on* the same line.

Such a set is called a *finit geometry*. It exemplifies the stark contrast with Euclidean geometry. *Point*, *line*, and the *on* relation are primitives specified by the above axioms only. Additional properties can be introduced to enrich the geometry.

2.3.1 Seven-Point Geometry

From the third axiom there are $n \geq 3$ points on a line. The smallest geometry, with $n = 3$ points on each line and a total of $n^2 - n + 1 = 7$ (see Exercise 3 below) points or lines is illustrated in Fig. 2.14. The lines are $a = [A, D, B]$, $b = [A, E, C]$, $c = [C, F, B]$, $d = [A, G, F]$, $e = [B, G, E]$, $f = [C, G, D]$, $g = [D, E, F]$. Observe that the projective plane axioms are satisfied (i.e., two distinct points are on a unique line and vice versa) for \forall lines $x \in \{a, \dots, g\}$

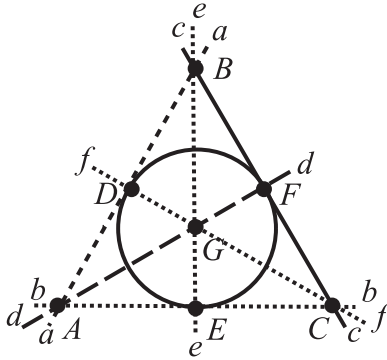


Figure 2.14. A “picture” of the 7-point geometry.

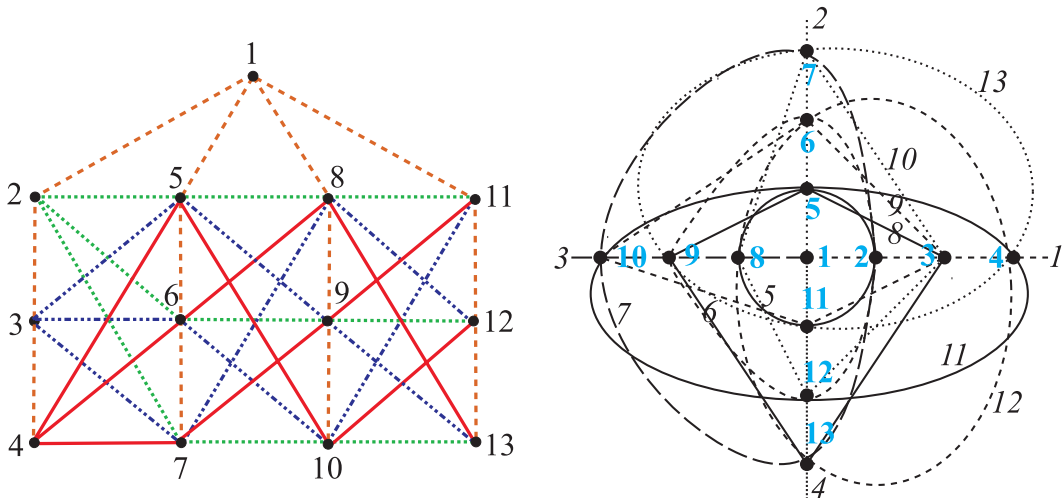


Figure 2.15. “Pictures” for 13-point geometry. There are four “lines” on each point and vice versa. On the right, the points are indicated in **bold numbers** and lines in *italics*.

and \forall points $X \in \{A, \dots, G\}$. The lines/curves joining the \bullet in the picture have no significance other than showing the association of the points belonging to a particular line.

↓ lines	1	2	3	4	5	6	7	8	9	10	11	12	13	← points
1	1	2	3	4										
2	1				5	6	7							
3	1							8	9	10				
4	1										11	12	13	
5		2			5			8			11			
6		2				6			9			12		
7		2					7			10			13	
8			3		5				9				13	
9			3			6				10	11			
10			3				7	8				12		
11				4	5					10		12		
12				4		6		8					13	
13				4			7		9		11			

Figure 2.16. Incidence matrix for the 13-point geometry.

2.3.2 Finite Geometry with 13 Points

Finite geometries can be pictured in different ways, as in Figure 2.15, and specified by an incidence matrix for points and lines as for the 13-point geometry. Here, it is more convenient to denote the points and lines by numbers rather than letters. Those interested in pursuing the fascinating subject of finite geometries are referred to [116] and [117].

Exercises

1. Prove the 3-D dual shown in Fig. 2.12, i.e., that three noncollinear planes are on a point, using only the projective 3-D space axioms.
2. The second axiom may be weakened to 2a: “If p and q are distinct lines, there is *at least* one point on *on* both lines.” Prove that with axiom 2 replaced by 2a above, the resulting set of points and lines also satisfies the projective plane axioms.
3. Prove that in a finite geometry with n points on any line there are $n^2 - n + 1$ points (or lines) altogether.
4. Construct diagrams for the geometries with $n = 5$ and 6 points and try to generalize. This is hard: you may want to look up a topic in combinatorics called “Latin squares.”

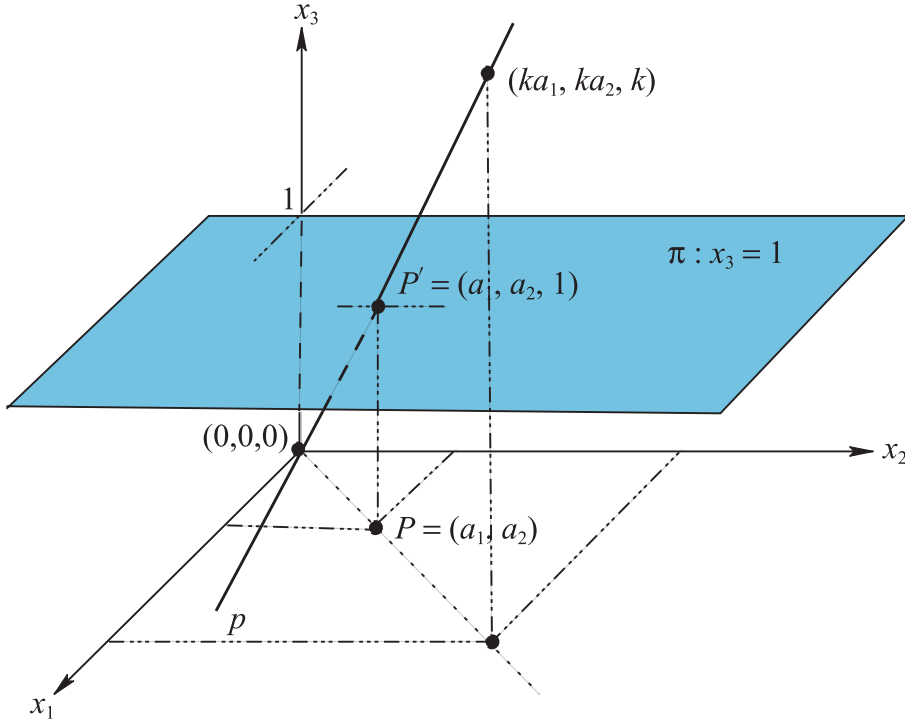


Figure 2.17. Homogeneous coordinates. Correspondence between points P in the x_1x_2 -plane and lines p through the origin in 3-D.

2.4 ♣ FT-2 Analytic Projective Geometry

Even though there is no notion of distance in projective space, it is possible to introduce a very useful coordinate system. The geometric development here is for the plane and can be generalized to projective N -space \mathbb{P}^N .

We consider the x_1x_2 plane in \mathbb{R}^3 with Cartesian coordinates as shown in Fig. 2.17 and the plane $\pi : x_3 = 1$. The point $P = (a_1, a_2)$ in the x_1x_2 -plane is identified with the point $P' = (a_1, a_2, 1)$ in the plane π . The line p through the points $(0, 0, 0)$ and P' is constructed and is described by the equations

$$p : \begin{cases} x_2 = \frac{a_2}{a_1}x_1, \\ x_3 = \frac{1}{a_2}x_2. \end{cases}$$

In this way, the point-to-line correspondence

$$P = (a_1, a_2) \in x_1x_2\text{-plane}$$

$$\longleftrightarrow p \subset x_1x_2x_3\text{-space, } p \text{ on } (a_1, a_2, 1), (0, 0, 0),$$

is established. Since p can be identified by any one of its points (ka_1, ka_2, k) for $k \neq 0$, the correspondence can be alternatively written as

$$P : (a_1, a_2) \longleftrightarrow (ka_1, ka_2, k), \quad 0 \neq k \in \mathbb{R}.$$

That is, (k_1a_1, k_1a_2, k_1) and any other multiple (k_2a_1, k_2a_2, k_2) for $k_1 \neq k_2$ represent the *same* point, since they both lie on the *same line* p . These triples are the *point coordinates* of P .

Let us return to Fig. 2.17 and imagine the point P moving farther away from the origin but always in the direction with slope $m = \frac{a_2}{a_1}$ approaching the ideal point in this direction. The point coordinates of P are $(a_1, a_2, 1)$ or $(1, m, \frac{1}{a_1})$, and as $a_1 \rightarrow \infty$ the ideal point is represented by the triples $(ka_1, ka_2, 0)$ for $k \neq 0$. The important realization is that the 0 in the third component identifies ideal points. This system is called *homogeneous coordinates*, allowing the representation both ordinary *and* ideal points, something that cannot be done with Cartesian coordinates; it has other nice properties as well.

Appealing to duality, a *line* x can be represented as an ordered triple of real numbers that is distinguished from a triple representing *points* using capitals for the coordinates themselves and “[]” instead of “()” as

$$x : [X_1, X_2, X_3] \neq [0, 0, 0],$$

again with the convention that $[kX_1, kX_2, kX_3]$, $\forall k \neq 0$, represents the same line. These triples are called *line coordinates*. So $(1, 2, 3)$ represents a point and $[1, 2, 3]$ a line. Actually, this is quite reasonable, for consider a line described in Cartesian coordinates by

$$A_1x_1 + A_2x_2 + A_3 = 0. \tag{2.1}$$

For $A_3 \neq 0$ the same line is also given by

$$\frac{A_1}{A_3}x_1 + \frac{A_2}{A_3}x_2 + 1 = 0. \tag{2.2}$$

In either case, the three constants A_1, A_2, A_3 completely specify the line. So taking the triple $[A_1, A_2, A_3]$ with the understanding that A_i is the coefficient of x_i for

$i = 1, 2$ and A_3 is the constant completely describes the line in (2.1). Equivalently, the triple $[\frac{A_1}{A_3}, \frac{A_2}{A_3}, 1]$ describes the line as given by (2.2), and hence these two triples and their multiples identify the same line.

There remains to define the *on* relation. We say that a point $X = (x_1, x_2, x_3)$ is *on* the line $x = [X_1, X_2, X_3]$, and dually the line x is *on* the point if and only if

$$X \cdot x = x \cdot X = x_1 X_1 + x_2 X_2 + x_3 X_3 = 0.$$

Consider now the three points $Y = (y_1, y_2, y_3)$, $Z = (z_1, z_2, z_3)$, $W = (w_1, w_2, w_3)$, all on the line $x = [X_1, X_2, X_3]$. Then

$$y_1 X_1 + y_2 X_2 + y_3 X_3 = 0,$$

$$z_1 X_1 + z_2 X_2 + z_3 X_3 = 0,$$

$$w_1 X_1 + w_2 X_2 + w_3 X_3 = 0.$$


This system of linear homogeneous equations has a nontrivial solution $[X_1, X_2, X_3]$ if and only if the determinant of the coefficients is zero. That is, Y, Z, W are on a line x if and only if

$$\begin{vmatrix} y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = 0. \quad (2.3)$$

This condition provides the line coordinates of x as

$$\left[\begin{vmatrix} y_2 & y_3 \\ z_2 & z_3 \end{vmatrix}, \begin{vmatrix} y_3 & y_1 \\ z_3 & z_1 \end{vmatrix}, \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \right]. \quad (2.4)$$

A direct consequence is that the points X, Y , and $\lambda X + \mu Y$ for $\lambda, \mu \in \mathbb{R}$ are collinear.

What are the line coordinates of the ideal line? We can find out by substituting two distinct ideal points in (2.4), i.e., $z_3 = y_3 = 0$. This yields $[0, 0, y_1 z_2 - y_2 z_1]$ or $[0, 0, 1]$. So far, it is all falling into place: we have a coordinate system that provides coordinates for every regular point (from which its Cartesian coordinates can be recovered) and *also* coordinates for the ideal elements (points and line). In addition, homogeneous coordinates can be extended to projective N -space and finite geometries.  **FT-2e**

Consider now the 7-point geometry (i.e., $n = 3$) described earlier and take the algebra J_2 , the integers modulo 2, with elements $\{1, 0\}$. The binary operations $+$ and \times are defined by

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} \times & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}.$$

Now we collect all triples (x_1, x_2, x_3) except $(0, 0, 0)$ and label them:

$$\begin{aligned} A &= (0, 0, 1), & B &= (0, 1, 0), & C &= (1, 0, 0), \\ D &= (0, 1, 1), & E &= (1, 0, 1), & F &= (1, 1, 0), & G &= (1, 1, 1). \end{aligned}$$

A pair of points is chosen specifying a line $x = [X_1, X_2, X_3]$ whose coordinates are computed from (2.4) for $X_i \in J_2$, yielding

$$\begin{aligned} a &= [1, 0, 0] & b &= [0, 1, 0] & c &= [0, 0, 1] \\ d &= [1, 1, 0] & e &= [1, 0, 1] & f &= [0, 1, 1] & g &= [1, 1, 1]. \end{aligned}$$

Note that each point is on two lines; for example A is *on* the lines having $X_3 = 0$. Similarly, other finite geometries can be coordinatized using the appropriate algebra J_p of integers modulo p , that is, the set of integers $\{1, 2, \dots, p-1\}$ with addition and multiplication analogously defined (see a book on modern algebra). To clarify, let us use J_3 to coordinatize the 13-point geometry. Excluding $(0, 0, 0)$, there are 26 triples (a_1, a_2, a_3) , $a_i \in J_3$, with the multiples $k(a_1, a_2, a_3)$, $k \neq 0 \in J_3$, also representing the same point. The only multiple of interest is $2(a_1, a_2, a_3)$, which equals (a_1, a_2, a_3) for $a_i \in J_3$, yielding 13 distinct triples. Similarly, with J_4 there are 63 nonzero triples whose multiples by 2 or 3 are the same in that algebra, yielding 21 distinct triples, which can therefore serve as the homogeneous coordinates of the 21-point geometry (Exercise 3 below).

2.4.1 Homogeneous Coordinates on a Line

Let us specialize homogeneous coordinates to a line o , which for convenience we take as being on the two points

$$O = (0, 1, 1), \quad U = (1, 1, 1).$$

The coordinates of a point $X = (x_1, x_2, x_3) \in o$ together with those of O and U satisfy the collinearity condition of (2.3); therefore $x_2 = x_3$, so $X = (x_1/x_2, 1, 1)$, with $X = (x_1, 0, 0)$ being the line's ideal point if $x_2 = 0$. Since there is no variation in the third component, the coordinates of regular points can be simplified to $O = (0, 1)$, $U = (1, 1)$ and in general, $X = (x_1, x_2)$, as indicated in Fig. 2.18. We can write

$$X : (x_1, x_2) = a_2(1, 1) + a_1(0, 1) = a_2U + a_1O$$

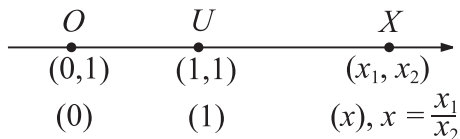


Figure 2.18. Homogeneous coordinates on the line.

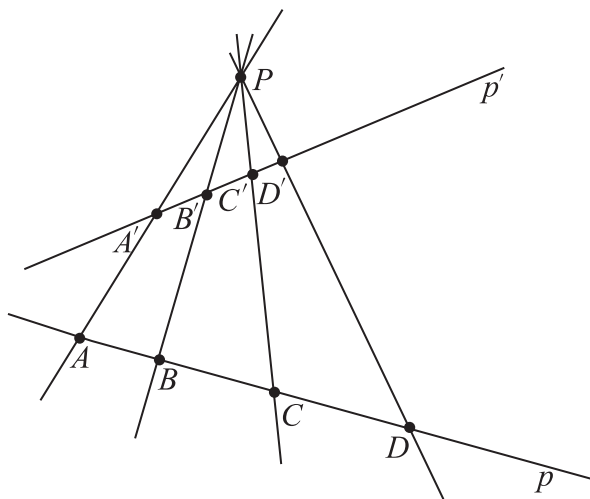


Figure 2.19. Two lines p, p' in perspective from the point P .

for $x_1 = a_2$, $x_2 = a_1 + a_2$, and O, U acting as a basis. The point O can be thought of as the “origin,” and U establishes the “unit”; recall that we do not have a metric in the projective plane. There is a further simplification obtained by dividing by x_2 to represent the point X by the single value $x = x_1/x_2$.

2.5 The Fundamental Transformations of Projective Geometry

A short diversion into the basic transformations of projective geometry is needed in order to prove a fundamental property of homogeneous coordinates. Consider two *directed* (in the Euclidean plane) lines p, p' and a point P not on the lines, as shown in Fig. 2.19. A line on P intersecting p at the point A and p' at A' establishes a one-to-one correspondence called *perspectivity* between the points of p and p' . It is denoted by

$$p(A, B, C, D, \dots) \stackrel{P}{=} p'(A', B', C', D', \dots).$$

The two lines are said to be in perspective with respect to the point P , which may be indicated above the $\stackrel{P}{=}$ symbol. Of course, there is the dual concept, shown in Fig. 2.20, of two points P, P' being in perspective with respect to a line p denoted by

$$P(a, b, c, d, \dots) \stackrel{\wedge}{=} P'(a', b', c', d', \dots).$$

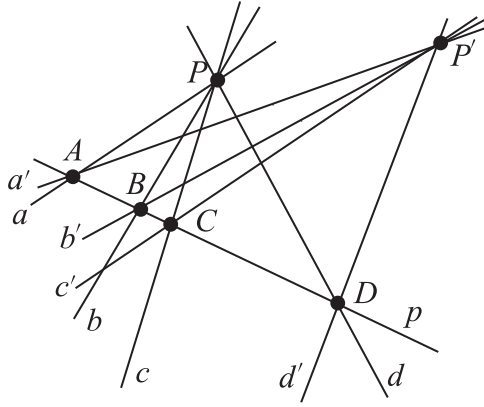


Figure 2.20. Two points P, P' in perspective from the line p .

Consider now a perspectivity $p(A, B, C, D, \dots) \stackrel{\wedge}{=} p'(A', B', C', D', \dots)$ as shown in Fig. 2.21 and the *cross-ratio*

$$(A, B; C, D) = \frac{AC/BC}{AD/BD}.$$

Theorem 2.5.1. *The cross-ratio is invariant under perspectivity.*

Proof. We give an elementary proof in the Euclidean plane, leaving the projective aspects for the exercises. With reference to Fig. 2.21,

$$(A, B; C, D) = \frac{\frac{IP \cdot AC}{IP \cdot BC}}{\frac{IP \cdot AD}{IP \cdot BD}} = \frac{\frac{\text{area} \triangle APC}{\text{area} \triangle BPC}}{\frac{\text{area} \triangle APD}{\text{area} \triangle BPD}} = \frac{\frac{AP \cdot CP \sin(\angle APC)}{BP \cdot CP \sin(\angle BPC)}}{\frac{AP \cdot DP \sin(\angle APD)}{BP \cdot DP \sin(\angle BPD)}},$$

or

$$(A, B; C, D) = \pm \frac{\sin(\angle APC) \cdot \sin(\angle BPD)}{\sin(\angle BPC) \cdot \sin(\angle APD)}. \quad (2.5)$$

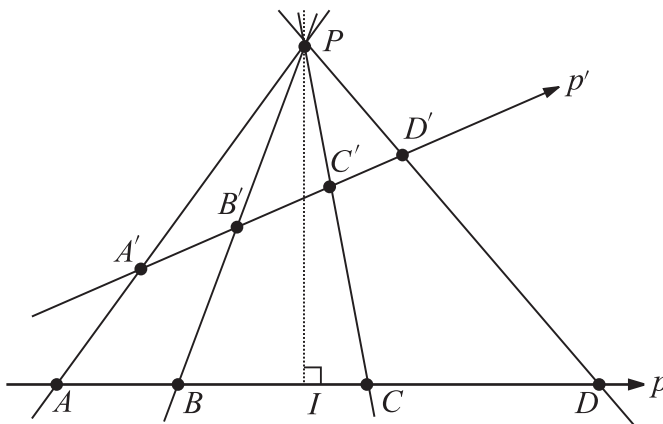


Figure 2.21. Diagram for the cross-ratio property.

Similarly, it can be shown that $(A', B'; C', D')$ also equals the right-hand-side of (2.5). There remains to show that the signs match (see exercises below). As stated, this definition of the cross-ratio applies only to the Euclidean plane, but it can be appropriately generalized to the augmented plane (π, p_∞) , the cross-section invariance being a *projective* property. ■

The set of all points on a line is called a *pencil of points*, and dually, the set of all lines on a point is called a *pencil of lines*. A pencil of points may undergo successive perspectivities, as in

$$p_1 \xrightarrow{P_1} p_2 \xrightarrow{P_2} \cdots \xrightarrow{P_{n-1}} p_n.$$

A sequence of perspectivities is called a *projectivity* and is denoted by $p \xrightarrow{\quad} p_n$. In projective geometry we study the properties that remain invariant under projectivities. We state without proof a couple of important results.

Theorem 2.5.2. *A projectivity is determined by the correspondence between two triples of distinct points. That is, if the triples (A, B, C) and (A', B', C') correspond under a projectivity $p \xrightarrow{\quad} p_n$, then the correspondence between any $X \in p$ and $X' \in p'$ is uniquely determined (i.e., projectivity is determined by just three points).*

This does not determine the actual sequence of perspectivities from which the projectivity resulted. Rather, it just determines the correspondence between the original and final pencils of points.

Theorem 2.5.3. *A projectivity can be expressed by at most two perspectivities.*

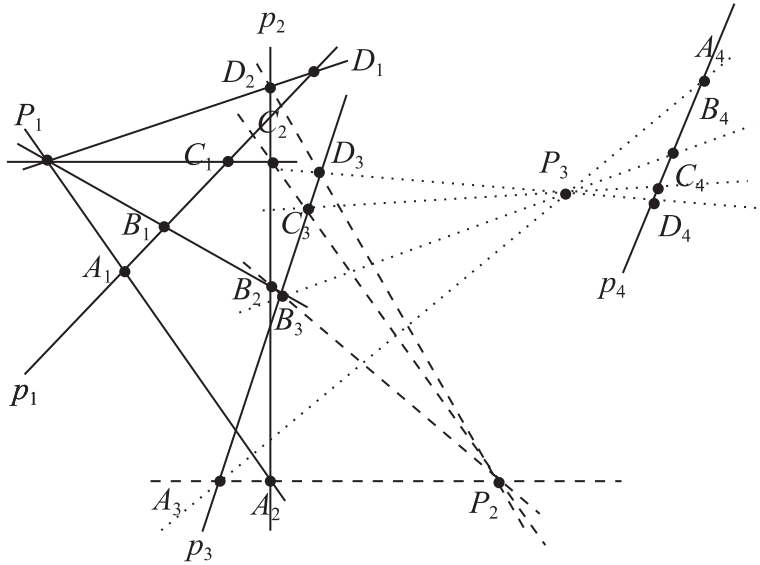


Figure 2.22. A projectivity composed of four perspectivities.

Exercises

1. Show that the 24 distinct permutations of four symbols A , B , C , and D result in exactly six different values of the cross-ratios as formed from these four symbols (corresponding to four points on a line)
2. Prove Theorem 2.5.1 when the perspectivity is with respect to an ideal point.
3. Prove Theorem 2.5.2 in the Euclidean plane and also with a perspectivity with respect to an ideal point.
4. Prove Theorem 2.5.3 in the Euclidean plane and also with a perspectivity with respect to an ideal point.

2.6 More on Homogeneous Coordinates

2.6.1 Linearizability of Projective Transformations

It is easy to see that the cross-ratio remains invariant under a projectivity and in fact, completely characterizes (determines) the projectivity. This invariance leads to a very important result and the reason homogeneous coordinates are so widely used in fields such as computer graphics. It is proved first for the plane \mathbb{P}^2 and then generalized to higher dimensions. For a projectivity $s \xrightarrow{\sim} s'$ between two directed lines

s and s' (as in Fig. 2.21, see lines p and p'), four points and their corresponding single homogeneous coordinates indicated by $()$, $P : (p)$, $Q : (q)$, $R : (r)$, $X : (x)$ are chosen to form the cross-ratio

$$(P, Q; R, X) = \frac{(p-r)(q-x)}{(p-x)(q-r)}.$$

Similarly, the cross-ratio

$$(P', Q'; R', X') = \frac{(p'-r')(q'-x')}{(p'-x')(q'-r')}$$

is formed with the corresponding points $P' : (p')$, $Q' : (q')$, $R' : (r')$, $X' : (x')$ on s' . By the cross-ratio invariance, under a projectivity, we have

$$(P, Q; R, X) = (P', Q'; R', X'), \quad (2.6)$$

yielding

$$x' = \frac{ax + b}{cx + d}. \quad (2.7)$$

The a, b, c, d are given in terms of the coordinates p, q, r, p', q' , and r' , and their actual values are immaterial to the development. This is a *linear fractional transformation*, also called a *Möbius transformation*, completely specifying the projectivity $s \xrightarrow{\quad} s'$. Reverting back to the ratio obtained from the homogeneous coordinates, i.e., $x = x_1/x_2$ and $x' = x'_1/x'_2$, we obtain from (2.7),

$$\frac{x'_1}{x'_2} = \frac{ax_1 + bx_2}{cx_1 + dx_2}. \quad (2.8)$$

Then rewriting the homogeneous coordinates of a point on a line in terms of two components as $X = (x_1, x_2)$, see Fig. 2.18, we get

$$k \left(\frac{x'_1}{x'_2}, 1 \right) = \left(\frac{ax_1 + bx_2}{cx_1 + dx_2}, 1 \right).$$

Here k is an arbitrary proportionality constant, which is needed, since *multiples* of a triple (the 1 in the third position is not shown; see Fig. 2.18) represent the same point. Eliminating the fractions yields

$$k(x'_1, x'_2) = (ax_1 + bx_2, cx_1 + dx_2),$$

from which we obtain the *linear* relations

$$kx'_1 = ax_1 + bx_2, kx'_2 = cx_1 + dx_2 \quad (2.9)$$

between the coordinates on the directed lines s and s' . The transformation (2.8) is called *projective*, and this is the most general class of *nonlinear* transformations that can be linearized by homogeneous coordinates. It is worth examining this in greater detail. Rewrite the transformation as

$$T(x) = \frac{ax + b}{cx + d} = \left(\frac{bc - ad}{c^2} \right) \cdot \left(\frac{1}{x + d/c} \right) + \left(\frac{a}{c} \right),$$

which provides the decomposition of T into

$$T = T_4 \circ T_3 \circ T_2 \circ T_1,$$

where \circ denotes the transformation composition and

$$T_1(x) = x + \frac{d}{c}, \quad T_2(x) = \frac{1}{x}, \quad T_3(x) = \frac{bc - ad}{c^2}x, \quad T_4(x) = x + \frac{a}{c}.$$

That is, T is composed of two translations T_1, T_4 , a scaling T_3 , and an inversion T_2 . Inversions and translations, being nonlinear, are linearized, and this property makes homogeneous coordinates especially useful in applications, since the *projective* transformations, translation, perspective, rotation, and scaling, can all be described in terms of matrices. Hence, a sequence of such transformations can be expressed as a matrix product and computed in a convenient and efficient way (e.g., see [146]).

For n -dimensional projective space \mathbb{P}^N and projectivities there, the linear fractional transformation (2.7) generalizes to

$$x_i'' = \frac{a_{1i}x_1' + a_{2i}x_2' + \cdots + a_{N,i}x_N' + a_{(N+1),i}}{b_{1i}x_1' + b_{2i}x_2' + \cdots + b_{N,i}x_N' + b_{(N+1),i}}, \quad i = 1, 2, \dots, n, \quad (2.10)$$

where we have reverted to homogeneous coordinates by

$$x_i'' = \frac{x_i'}{x_{(N+1)}'}, \quad x_j' = \frac{x_j}{x_{(N+1)}}, \quad j = 1, 2, \dots, (n+1).$$

Here the x_i are Cartesian coordinates, while the x_i' and x_i'' are homogeneous coordinates.

2.7 ** The Linearized Projective Transformations

To recap, the projective transformation

$$T(x) = ax \quad (2.11)$$

is linear and is called *scaling* when $a > 0$, $a \neq 1$; then it is a *contraction* for $a < 1$ and a *stretching*, or *magnification*, when $a > 1$. For complex $a = r(\cos \theta + i \sin \theta)$, the result is a *rotation* of x by an angle θ about the origin as well as a scaling by r . By contrast, *translation*

$$T(x) = x + a \quad (2.12)$$

and *inversion or perspective*

$$T(x) = \frac{1}{x} \quad (2.13)$$

are not linear, and for these the linearization, accomplished via homogeneous coordinates, enables the transformations to be expressed in matrix form. This is of importance in such applications as *computer graphics* and *geometric modeling*.

2.7.1 Projective Transformations in the Plane \mathbb{P}^2

For notational convenience, a point $P = (x_1, x_2, x_3)$ and its transform $P' = (x'_1, x'_2, x'_3)$, expressed in homogeneous coordinates, are taken as row vectors, while the transformations are denoted by two capital letters. We take $x'_3 = c'$ and $x_3 = c$, such that for ordinary points, $c, c' \neq 0$, and $c, c' = 0$ for ideal points. Many references in computer graphics, geometric modeling, robotics, and other fields consider only the case $x_3 = 1$, which unnecessarily excludes the ideal points.

The scaling transformation $P' = PS(s_1, s_2)$, where $x'_i = s_i x_i$ and $x'_3 = x_3 = c$, in matrix form is

$$PS: \quad (x'_1, x'_2, c) = (x_1, x_2, c) \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (2.14)$$

The rotation $P' = PR(\theta)$ by an angle θ about the origin is

$$PR: \quad (x'_1, x'_2, c) = (x_1, x_2, c) \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (2.15)$$

For the translation $P' = PT(t_1, t_2)$, with $x'_i = x_i + t_i$, we can set $c = 1$, since translation of ideal points is not meaningful, to obtain

$$PT: \quad (x'_1, x'_2, 1) = (x_1, x_2, 1) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_1 & t_2 & 1 \end{pmatrix}. \quad (2.16)$$

Successive scaling or rotation transformations result in successive multiplications of the corresponding matrices, as they should, since these are linear transformations. The fun and beauty is that it also works with translations. For $P' = PT_1(t_1^1, t_2^1)$ and $P'' = P'T_2(t_1^2, t_2^2)$, the result is $P'' = PT_1T_2 = PT_2T_1$ so that $x_i'' = x_i + t_i^1 + t_i^2$. The matrices commute with respect to multiplication, as they should, since the result of successive translations does not depend on their order.

The composition of scalings, rotations, and translations is obtained by the multiplication of the corresponding matrices. In addition, the matrices are sparse (i.e., they have many zeros), and the computational advantage is tremendous.

2.7.2 Projective Transformations in \mathbb{P}^3

The generalizations of these transformations to 3-D are straightforward, with the previous remarks regarding $x_3 = c$ now pertaining to $x_4 = c$. For scaling, $P' = PS(s_1, s_2, s_3)$,

$$PS: (x'_1, x'_2, x'_3, c) = (x_1, x_2, x_3, c) \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (2.17)$$

Let the matrices for the rotation about the x_i -axes by an angle θ be denoted by $PR_i(\theta)$, $i = 1, 2, 3$. The rotation about x_3 is given by

$$PR_3: (x'_1, x'_2, x'_3, c) = (x_1, x_2, x_3, c) \begin{pmatrix} \cos \theta & \sin \theta & 0 & 0 \\ -\sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (2.18)$$

and is a direct extension of the 2-D rotation. The two other rotation matrices are

$$PR_1(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (2.19)$$

$$PR_2(\theta) = \begin{pmatrix} \cos \theta & 0 & -\sin \theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \theta & 0 & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

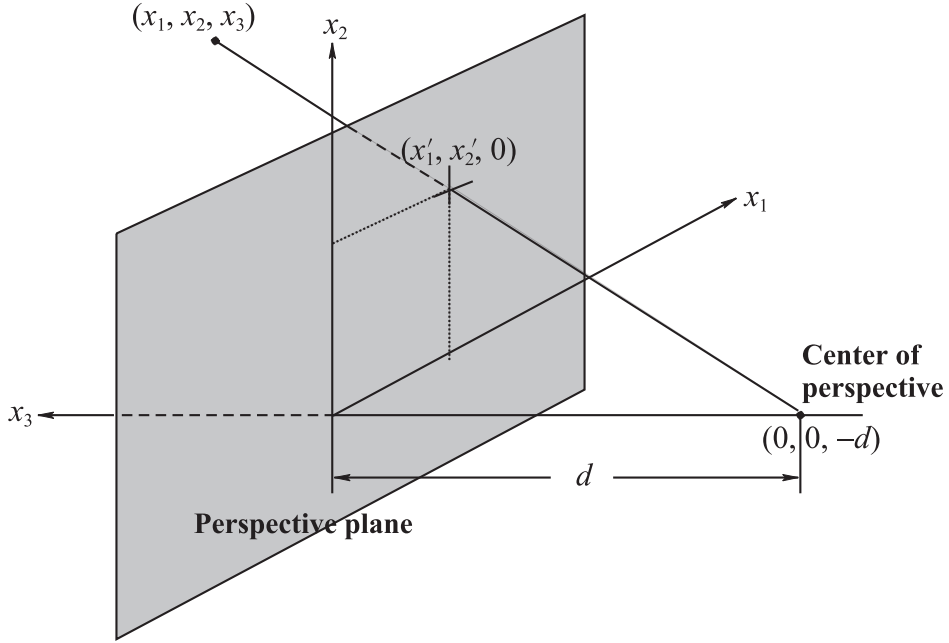


Figure 2.23. Perspective with respect to the x_3 -axis.

For 3-D translations $P' = PT(t_1, t_2, t_3)$, $c = 1$ as in 2-D, and

$$PT: \quad (x'_1, x'_2, x'_3, 1) = (x_1, x_2, x_3, 1) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ t_1 & t_2 & t_3 & 1 \end{pmatrix}. \quad (2.20)$$

In 3-D, the perspective transformation is important, for it provides images with a sense of depth and realism. It is convenient to choose the center of the perspective transformation on the negative x_3 -axis at a distance d behind the perspective plane, as shown in Fig. 2.23. The transformation equations, obtained by similar triangles, are

$$x'_1 = \frac{x_1}{x_3/d + 1}, \quad x'_2 = \frac{x_2}{x_3/d + 1}, \quad x'_3 = 0. \quad (2.21)$$

Denote the perspective transformation matrix by $V[\pi_{per}, C_{per}]$, where π_{per} and C_{per} are the perspective plane and center respectively. For the setup in Fig. 2.23, $P' = PV[x_3 = 0, (0, 0, -d)]$, and explicitly,

$$(x'_1, x'_2, 0, w) = (x_1, x_2, x_3, c) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/d \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (2.22)$$

with the relation between c and w emerging next. The perspective image of an ideal point $P = (1, x_2/x_1, x_3/x_1, 0)$, called a *vanishing point*, is the *ordinary point* $P' = (x'_1, x'_2, 0, x_3/d) = (dx_1/x_3, dx_2/x_3, 0, 1)$. Hence

$$w = \frac{x_3}{d} + c, \quad x'_1 = \frac{x_1}{w}, \quad x'_2 = \frac{x_2}{w}, \quad x'_3 = 0.$$

All projective transformations can now be performed successively by matrix multiplication. The matrix of the perspective transformation, unlike those of other projective transformations, is singular, confirming that an object cannot be uniquely retrieved from its perspective image.

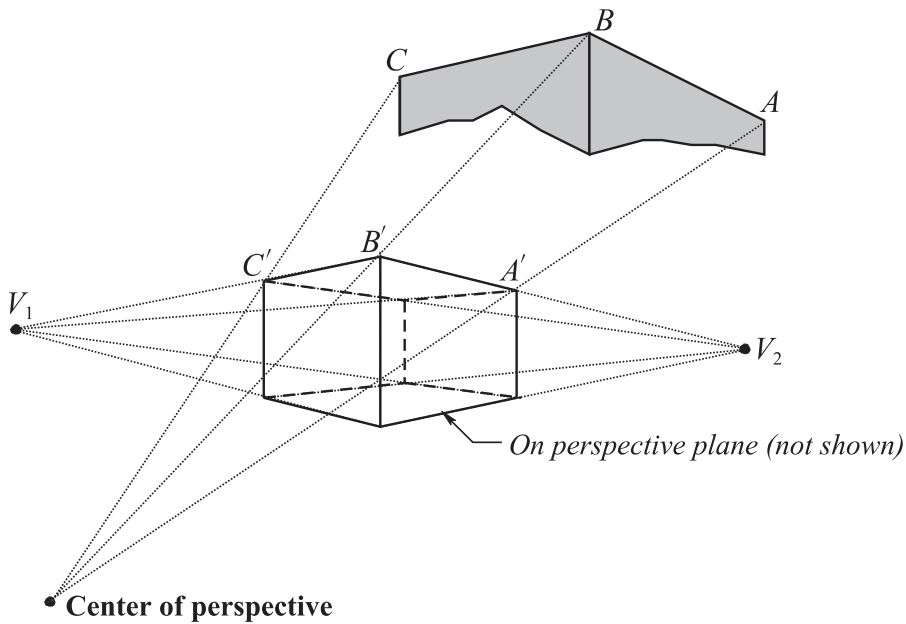


Figure 2.24. The perspective image of an object (partially shown).

In the figure, the perspective transformation of three vertices A, B, C is traced, showing their images A', B', C' . This is sometimes called a *2-point perspective*, having two *vanishing points* V_1 and V_2 , the images of ideal points, giving the image “3-dimensionality.” Actually, the ideal point in the vertical direction is the “third vanishing point.” It would have been a *3-point perspective* had the perspective plane not been parallel to the front (vertical) face of the object. All lines joining the object’s points with their images intersect at the *center of perspective*.

2.8 ** Coordinate System Rotations

This section, complementing the discussion of rotations about one axis above, is adapted from an old gem [161, pp. 240–241] (the textbook from which I learned analytic geometry). In \mathbb{R}^3 with a positive direction assigned to a line ℓ , translate the coordinate system to any point $Q \in \ell$, denoting by α, β, γ the *direction angles* between the positive direction of ℓ and the translates of the axes x_1, x_2, x_3 respectively. Their cosines, called *direction cosines*, have the property

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1. \quad (2.23)$$

Denoting by $\hat{\mathbf{e}}_i$ the unit vectors in the x_i directions, we obtain that

$$\hat{\ell} = \hat{\mathbf{e}}_1 \cos \alpha + \hat{\mathbf{e}}_2 \cos \beta + \hat{\mathbf{e}}_3 \cos \gamma \quad (2.24)$$

is the unit vector in the positive direction of ℓ , with (2.23) equivalent to $\hat{\ell} \cdot \hat{\ell} = |\hat{\ell}|^2$. The cosine of the angle ϕ between two directed lines ℓ_1, ℓ_2 is given in terms of their corresponding unit vectors $\hat{\ell}_1, \hat{\ell}_2$, and their corresponding direction angles, subscripted by 1 and 2, respectively, are given by

$$\cos \phi = \hat{\ell}_1 \cdot \hat{\ell}_2 = \cos \alpha_1 \cos \alpha_2 + \cos \beta_1 \cos \beta_2 + \cos \gamma_1 \cos \gamma_2, \quad (2.25)$$

with $\cos \phi = 0$ when the lines are perpendicular.

For a plane π , let n be the directed line through the origin O and perpendicular to π with direction angles α, β, γ . The normal form of the equation of π is

$$\pi : x_1 \cos \alpha + x_2 \cos \beta + x_3 \cos \gamma = d, \quad (2.26)$$

with d the directed distance of π from O . It is an easy matter to verify that the directed distance p of π from the point $P = (p_1, p_2, p_3)$ is given by

$$p = p_1 \cos \alpha + p_2 \cos \beta + p_3 \cos \gamma - d. \quad (2.27)$$

With reference to Fig. 2.25, the coordinate system x_1, x_2, x_3 is rotated about the origin O to the system x'_1, x'_2, x'_3 , all points remaining stationary with a point P having two sets of coordinates $(x_1, x_2, x_3), (x'_1, x'_2, x'_3)$. Our task is to find the relationship between them. Letting the direction cosines of the x'_i axes with respect to x_i be $\cos \alpha_i, \cos \beta_i, \cos \gamma_i$ for $i = 1, 2, 3$ respectively, we obtain

$$\begin{aligned} \cos^2 \alpha_i + \cos^2 \beta_i + \cos^2 \gamma_i &= 1, \\ \cos \alpha_j \cos \alpha_k + \cos \beta_j \cos \beta_k + \cos \gamma_j \cos \gamma_k &= 0, \end{aligned} \quad (2.28)$$

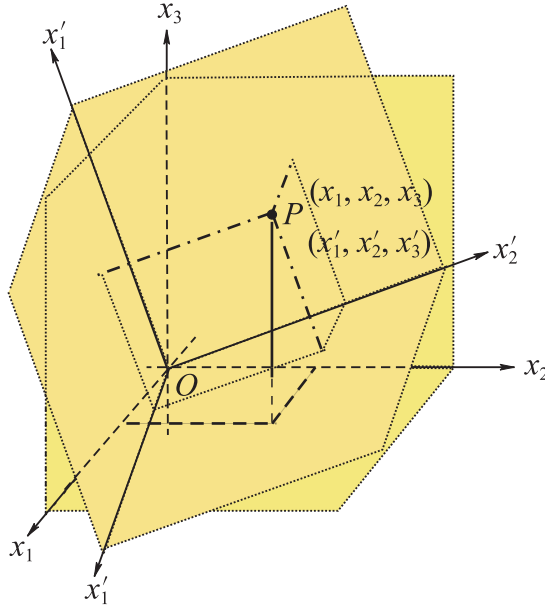


Figure 2.25. Rotation of the coordinate system $x_1x_2x_3$ about the origin.

where $j \neq k$, $j, k = 1, 2, 3$, the first part corresponding to (2.23) for the x'_i and the second due to the orthogonality between the pairs x'_j, x'_k of axes. The plane $x'_2x'_3$, which passes through the origin O , has equation

$$x'_2x'_3 : x_1 \cos \alpha_1 + x_2 \cos \beta_1 + x_3 \cos \gamma_1 = 0,$$

with coefficients the direction cosines of the axis x'_1 , which is orthogonal to it. Similarly, for all the 2-planes,

$$x'_jx'_k : x_1 \cos \alpha_i + x_2 \cos \beta_i + x_3 \cos \gamma_i = 0, \quad i \neq j \neq k. \quad (2.29)$$

The directed distance of the point P from the plane $x'_jx'_k$ is its x'_i coordinate, $i \neq j \neq k$; hence by (2.27), (2.29),

$$x'_i = x_1 \cos \alpha_i + x_2 \cos \beta_i + x_3 \cos \gamma_i. \quad (2.30)$$

These describe the coordinate transformation $\{x_1, x_2, x_3\} \rightarrow \{x'_1, x'_2, x'_3\}$. Reversing the roles of x_i and x'_i in the previous reasoning, we obtain the inverse transformation $\{x'_1, x'_2, x'_3\} \rightarrow \{x_1, x_2, x_3\}$:

$$x_i = x_1 \cos \alpha_i + x_2 \cos \beta_j + x_3 \cos \gamma_k, \quad i \neq j \neq k. \quad (2.31)$$

For consistency with Section 2.7, (2.30) and (2.31) are rewritten in homogeneous coordinates and in matrix form:

$$(x'_1, x'_2, x'_3, 1) = (x_1, x_2, x_3, 1) \begin{pmatrix} \cos \alpha_1 & \cos \beta_1 & \cos \gamma_1 & 0 \\ \cos \alpha_2 & \cos \beta_2 & \cos \gamma_2 & 0 \\ \cos \alpha_3 & \cos \beta_3 & \cos \gamma_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (2.32)$$

$$(x_1, x_2, x_3, 1) = (x'_1, x'_2, x'_3, 1) \begin{pmatrix} \cos \alpha_1 & \cos \alpha_2 & \cos \alpha_3 & 0 \\ \cos \beta_1 & \cos \beta_2 & \cos \beta_3 & 0 \\ \cos \gamma_1 & \cos \gamma_2 & \cos \gamma_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (2.33)$$

The matrix A of the transformation in (2.30) has determinant $|A| = 1$, and the matrix of the inverse transformation (2.31) is the transpose A^T . These, and the matrices corresponding to the rotations about one axis in Section 2.7, are examples of *orthogonal* matrices (transformations), which also appear in the previous section. As a reminder, the above equations *together* with the six conditions of (9.47) specify the rotation of the coordinate system about the origin. By the way, such a rotation about an arbitrary point can be obtained by translating the origin first and then rotating, or equivalently, multiplying the matrix A in (9.48) by that in (2.20), with resulting matrix describing the combined transformation. That is,

$$(x'_1, x'_2, x'_3, 1) = (x_1, x_2, x_3, 1) \begin{pmatrix} \cos \alpha_1 & \cos \beta_1 & \cos \gamma_1 & 0 \\ \cos \alpha_2 & \cos \beta_2 & \cos \gamma_2 & 0 \\ \cos \alpha_3 & \cos \beta_3 & \cos \gamma_3 & 0 \\ t_1 & t_2 & t_3 & 1 \end{pmatrix}. \quad (2.34)$$

From the vector and matrix form, the generalization to \mathbb{R}^N is direct, with a slight change in notation denoting the direction angle between the x_i axis and a directed line $\ell \subset \mathbb{R}^N$ by α_i and the unit vector in the direction of ℓ :

$$\hat{\ell} = \sum_{i=1}^N \hat{\mathbf{e}}_i \cos \alpha_i, \quad \sum_{i=1}^N \cos^2 \alpha_i = 1. \quad (2.35)$$

Proceeding as above, the corresponding $(N+1) \times (N+1)$ transformation matrix is obtained *together* with the $(N^2 + N)/2$ additional conditions analogous to (9.47): N conditions from the above equation and $N(N-1)/2$ more due to the orthogonality of all pairs of axes.

2.9 ** Analytic Proofs

Desargues's theorem will be helpful in subsequent chapters. This is an opportune time to prove it and learn yet another nice way to utilize homogeneous coordinates.

2.9.1 Computational Proof

Three noncollinear points P_i , specified in homogeneous coordinates by triples $A_i, i = 1, 2, 3$, are chosen on the projective plane \mathbb{P}^2 . Another point P , with triple A , is selected such that $A = c_1A_1 + c_2A_2 + c_3A_3$ with all $c_i \neq 0$. Each point P_i is also represented by the triple $c_iA_i = A_i^*$, which are fixed from now on, leaving $A = A_1^* + A_2^* + A_3^*$. Any other point Q can be described as $Q = a_1A_1^* + a_2A_2^* + a_3A_3^*$ with the A_i^* of the multiples now fixed. The coefficients can be used as a new set of coordinates, i.e., $Q = (a_1, a_2, a_3)$ and $P_1 = (1, 0, 0), P_2 = (0, 1, 0), P_3 = (0, 0, 1), P = (1, 1, 1)$. In this coordinate system it is easy to see that $Q = (a_1, a_2, a_3)$ and $Q' = k(a_1, a_2, a_3), k \neq 0$, represent the same point, allowing us to write $Q = Q'$. This is similar to homogeneous coordinates with 0 in the third position no longer implying that the point is ideal. This turns out to be a very useful tool.

To illustrate, we provide an analytic proof of Desargues's theorem. It is convenient to take the triangle's vertices A, B, C as the basis triples and $P = A + B + C$ or $P = (1, 1, 1)$ as the point of perspectivity, as shown in Fig. 2.26. The three lines from P through the vertices of the triangle are denoted by $a', b',$ and c' . With $A = (1, 0, 0)$ and $P = (1, 1, 1)$ on a' , from (2.4) we find that

$$a' = \left[\begin{vmatrix} 1 & 1 \\ 0 & 0 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} \right] = [0, 1, -1].$$

For a point $A' = (a, a_1, a_2) \in a'$, where $A \neq A'$, the collinearity of P, A, A' and (2.3) yield

$$\begin{vmatrix} a & a_1 & a_2 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{vmatrix} = 0,$$

which shows that $a_1 = a_2$. Since these values are arbitrary, we can take $A' = (a, 1, 1)$. Similarly, we obtain the line coordinates $b' = [1, 0, -1], c' = [1, -1, 0]$ and the point coordinates $B' = (1, b, 1), C' = (1, 1, c)$. Let r be the line determined by A and B . Then

$$r = \left[\begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}, \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}, \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \right] = [0, 0, 1].$$

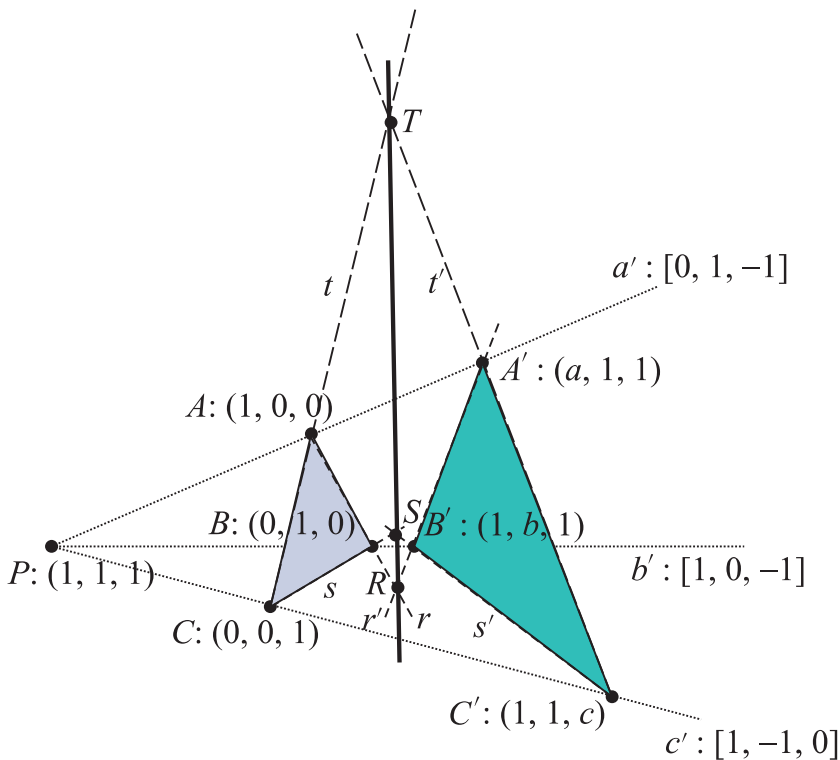


Figure 2.26. Diagram for analytic proof of Desargue's theorem.

Similarly, for the line r' determined by A' and B' ,

$$r' = \left[\begin{array}{c|c|c} 1 & 1 & \\ b & 1 & \end{array} , \begin{array}{c|c|c} a & 1 & \\ 1 & 1 & \end{array} , \begin{array}{c|c|c} a & 1 & \\ 1 & b & \end{array} \right] = [1-b, a-1, ab-1].$$

To find the intersection $r \cap r' = R$, we can use the *dual* of (2.4). That is, in its original form, (2.4) yields the coordinates of the *line* determined by the coordinates of two of its points. In the dual, the same equation yields the coordinates of the *point* on the two lines from their coordinates. Namely,

$$R = \left(\begin{vmatrix} 0 & 1 \\ a-1 & ab-1 \end{vmatrix}, \begin{vmatrix} 1 & 0 \\ ab-1 & 1-b \end{vmatrix}, \begin{vmatrix} 0 & 0 \\ 1-b & a-1 \end{vmatrix} \right) \\ = (1-a, b-1, 0).$$

Continuing in the same way, we obtain

$$S = BC \cdot B'C' = s \cap s' = (0, 1 - b, c - 1),$$

$$T = AC \cdot A'C' = t \cap t' = (a - 1, 0, 1 - c).$$

Applying again (2.3) with the coordinates found for R, S, T yields

$$\begin{vmatrix} 1 - a & b - 1 & 0 \\ 0 & 1 - b & c - 1 \\ a - 1 & 0 & 1 - c \end{vmatrix} = 0,$$

showing that R, S , and T are always collinear, proving Desargues's theorem in the *projective plane*.

2.9.2 Algebraic Proof

When the coordinates of the points and lines are not explicitly needed, their computation can be avoided with an algebraic proof. In Fig. 2.26, let all points be identified by their homogeneous coordinate triples. By the collinearity of P, A, A' there exist nonzero constants c_1, c_2 such that

$$P = c_1 A + c_2 A'. \quad (2.36)$$

The multiple $c_1 A$ identifies the *same* point as A , and similarly, $c_2 A'$ identifies $-A'$. Without loss of generality we can take the triples $c_1 A, c_2 A'$ for the representation of the points $A, -A'$, simplifying (2.37) to

$$P = A - A'. \quad (2.37)$$

The new triples for A, A' then remain *fixed* as in the previous section. Similarly, there exist constant multiples of B, B', C, C' such that

$$P = B - B', \quad P = C - C'. \quad (2.38)$$

Then

$$A - B = A' - B'. \quad (2.39)$$

The point $A - B$ is on the line AB , and the point $A' - B'$ is on $A'B'$. Therefore,

$$A - B = A' - B' = R = AB \cap A'B'. \quad (2.40)$$

Continuing, we obtain

$$C - A = C' - A' = T = AC \cap A'C', \quad B - C = B' - C' = S = BC \cap B'C'. \quad (2.41)$$

Altogether,

$$R + S + T = A - B + C - A + B - C = 0, \quad (2.42)$$

showing that the intersection points R , S , and T are collinear.

It is worth contrasting these proofs with the synthetic proofs of Desargues's theorem, as illustrated in Fig. 2.27.

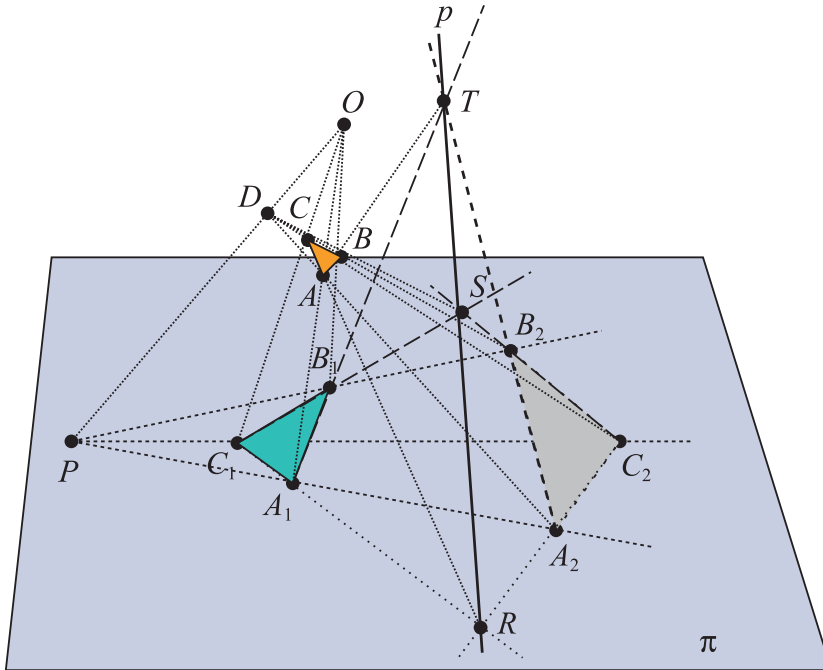


Figure 2.27. Diagram for the synthetic proof of Desargues theorem.

This proof involves embedding the 2-D perspectivity in 3-D. The theorem statement is that *two triangles in perspective from a point P are in perspective from a line p .*

Fun and Games with the ILM

Open the ILM: set *Chapter = Geometry* and click on “*Lock Guides.*”

The display shows **Desargues’s theorem** with the three intersection points R, S, T collinear.

Click and move any of the **red** triangle vertices along the lines.

The three points R, S, T remain collinear.

Next set *Example = Pappus* and click on “*Lock Guides.*”

The display shows **Pappus’s theorem**.

Again click on and move any of the six points A_i, B_i along the lines.

The three intersection points C_i remain collinear.

Exercises

1. Provide an analytic proof of Pappus’s theorem.
2. Generalize and find the representation of the ideal elements of (\mathbb{R}^3, p_∞) in terms of ordered quadruples.
3. Generalize the representation of the ideal elements of \mathbb{P}^N in terms of ordered $(N + 1)$ -tuples.

2.10 ** Conics on the Projective Plane

Curves can be considered as a collection of points and dually as a collection of lines whose *envelope* forms a curve, (see Fig. 2.28.) To distinguish them, they are referred to as *point-curves* and *line-curves* respectively. The duality of the projective plane enables us to treat them equivalently and leads to some nice results, illustrated here with conic point- and line-curves.

2.10.1 In the Projective Plane All Conics Are Ellipses

That all conics are ellipses in the projective plane can be observed on the right of Fig. 2.28, which relates to the position of an ellipse with respect to the ideal line l_∞ . That is, an ordinary (Euclidean) ellipse has no ideal points, and a Euclidean parabola corresponds to an ellipse tangent to the ideal line and hence having one ideal point corresponding to the direction in which the parabola “opens toward

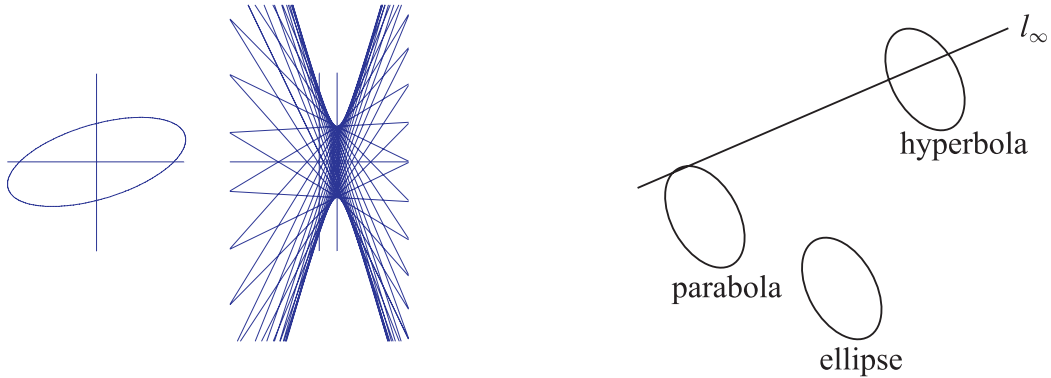


Figure 2.28. Ellipse (point-curve) and hyperbola (line-curve). Projective conics are ellipses.

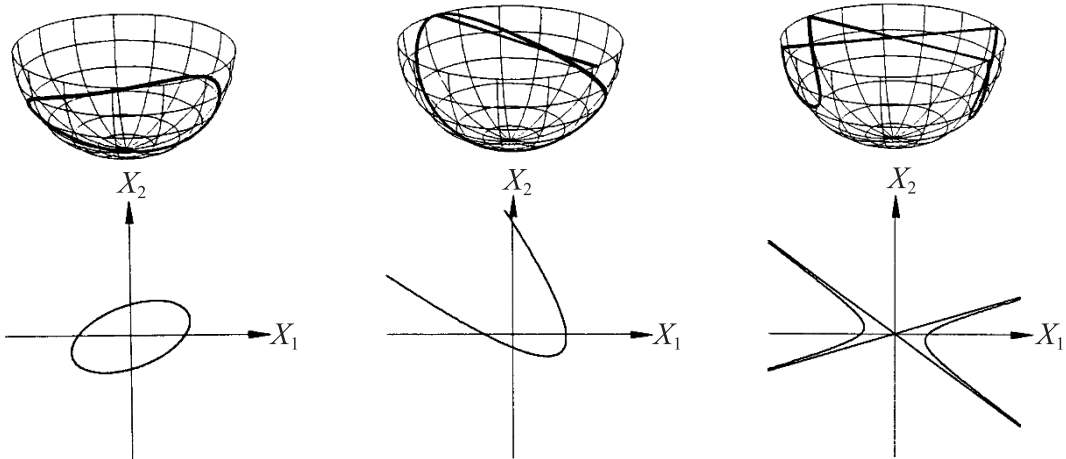


Figure 2.29. Ellipse (left), parabola (center), and hyperbola (right) in the projective plane model.

infinity.” A Euclidean hyperbola is a projective ellipse intersecting l_∞ at two ideal points corresponding to the directions of its asymptotes. The three cases in Fig. 2.29 are shown in the projective plane model.

2.10.2 Generalizing Pappus’s Theorem

In 1640, Pascal generalized Pappus’s theorem to conics.

Theorem 2.10.1 (Pascal) *If a simple hexagon $A_1A_2A_3B_1B_2B_3$ is inscribed in a conic, the intersections $C_i = A_jB_k \cdot A_kB_j$, $i \neq j \neq k$, are collinear.*

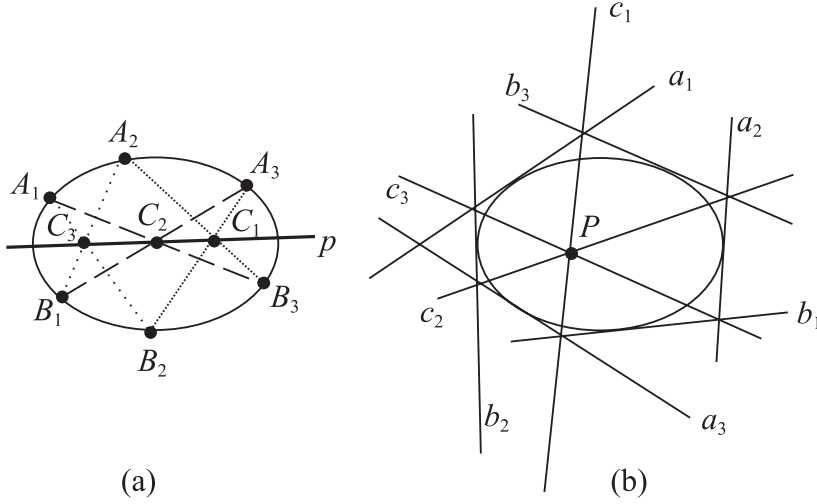


Figure 2.30. Pascal's theorem (a) and its dual, Brianchon's theorem (b).

Brianchon proved the result depicted in Fig. 2.30 in 1806, which turned out to be the *dual* of Pascal's theorem.

2.11 ** Geometries and Invariants

As already mentioned in connection with (2.10), homogeneous coordinates can be extended to higher dimensions. Specifically, a point $P \in \mathbb{P}^N$ is represented by an $(N+1)$ -tuple, the *point* coordinates

$$P = (x_1, x_2, \dots, x_N, x_{N+1}) \text{ or } (kx_1, kx_2, \dots, kx_N, kx_{N+1}), k \neq 0,$$

where not all components are zero. Dually, a hyperplane π is represented by the $(N+1)$ -tuple

$$\pi = [X_1, X_2, \dots, X_N, X_{N+1}], \text{ or } [kX_1, kX_2, \dots, kX_N, kX_{N+1}],$$

where again not all components are zero, these being the *hyperplane* coordinates of π . The *on* relation as before is defined by the inner product of the coordinates, i.e.,

$$P \text{ on } \pi \iff \sum_{i=1}^{N+1} x_i X_i = 0.$$

A projective transformation in \mathbb{P}^N is expressed by (2.10) and can be linearized by homogeneous coordinates. Projective transformations leave the cross-ratio invariant, and they are the most general transformations employed in projective geometry.

Various geometries were developed during the nineteenth century in addition to the Euclidean: affine, projective, spherical and inversive. Felix Klein proposed a way of classifying geometries that became known as the Erlangen Program, since he was a professor at the University of Erlangen. A geometry is regarded as a *space* together with a *group of transformations* acting on it. Geometrical properties are those remaining *invariant* under the group of transformations. As a reminder, a group is a set of elements G and a binary operation \circ with

$$\begin{aligned} a \circ b &= c \in G \quad \forall a, b \in G \text{ (closure property),} \\ (a \circ b) \circ c &= a \circ (b \circ c) \text{ (associativity)} \\ \exists e \in G \ni a \circ e &= e \circ a = a, \quad \forall a \in G \text{ (identity)} \\ \forall a \in G \exists a^{-1} \ni a^{-1} \circ a &= a \circ a^{-1} = e \text{ (inverse).} \end{aligned}$$

Exercises

Show that the set of all linear fractional transformations (i.e., the projective transformations) with the composition operation form a group.

The group of projective transformations has important special cases that form subgroups of transformations, i.e., subsets that are groups. With slight changes in the notation, consider the case of *translations*, for which the b 's in the denominator of (2.10) are zero except for the last one, i.e.,

$$\mathbf{x}' = \mathbf{x}A + \mathbf{c},$$

with row vectors denoted in **bold**, e.g., \mathbf{x} , and A is an $N \times N$ matrix. In general, a mapping

$$L : D \rightarrow D$$

is a *linear transformation* if it preserves linear combinations, i.e.,

$$L(a\mathbf{x} + b\mathbf{y}) = aL(\mathbf{x}) + bL(\mathbf{y}).$$

A general *affin* transformation can be written as

$$T(\mathbf{u}) = L(\mathbf{u}) + \mathbf{c},$$

where \mathbf{c} is a constant vector and L is a linear transformation. In this case,

$$T(a\mathbf{x} + b\mathbf{y}) = aL(\mathbf{x}) + bL(\mathbf{y}) + \mathbf{c} \neq aT(\mathbf{x}) + bT(\mathbf{y}) = aL(\mathbf{x}) + bL(\mathbf{y}) + 2\mathbf{c}.$$

In short, translations are not linear transformations. Affine transformations leave *convex* combinations ($\sum a_i x_i$, where $\sum a_i = 1$) invariant and also preserve *incidence* (i.e., the *on* relation).

Let us return to the representation

$$\mathbf{x}' = \mathbf{x}A + \mathbf{c}$$

and consider the case where the matrix has the property $AA^T = \lambda^2 I$, where λ is real and T denotes the transpose of a matrix. This is the *similarity* group. First take the case $\mathbf{c} = \mathbf{0}$. The square of the length, $|\mathbf{x}|$, of \mathbf{x} is given by $\mathbf{x} \cdot \mathbf{x} = \mathbf{x}\mathbf{x}^T$ and is called the *norm* of \mathbf{x} . For the similarity transformation $\mathbf{x}' = \mathbf{x}A$, we have

$$\mathbf{x}' \cdot \mathbf{x}' = (\mathbf{x}A) \cdot (\mathbf{x}A) = \lambda^2 \mathbf{x} \cdot \mathbf{x}.$$

Effectively, $|\mathbf{x}|$ has been amplified by $\pm\lambda$. For two vectors \mathbf{x}, \mathbf{y} , under similarity we obtain

$$\mathbf{x}' \cdot \mathbf{y}' = \lambda^2 \mathbf{x} \cdot \mathbf{y} \quad \mathbf{x} \cdot \mathbf{y} = |\mathbf{x}| \cdot |\mathbf{y}| \cos(\mathbf{x}, \mathbf{y}),$$

where (\mathbf{x}, \mathbf{y}) is the angle between the two vectors. So

$$|\mathbf{x}'| \cdot |\mathbf{y}'| \cos(\mathbf{x}', \mathbf{y}') = \lambda^2 |\mathbf{x}| \cdot |\mathbf{y}| \cos(\mathbf{x}, \mathbf{y}).$$

Therefore, $\mathbf{x} \perp \mathbf{y} \iff \mathbf{x}' \perp \mathbf{y}'$, and in general, $\cos(\mathbf{x}', \mathbf{y}') = \cos(\mathbf{x}, \mathbf{y})$. So under similarity, the ratio of distances and angles between vectors is preserved, or

$$\frac{|\mathbf{x}'|}{|\mathbf{y}'|} = \frac{|\mathbf{x}|}{|\mathbf{y}|}.$$

Specializing further to *orthogonal* matrices with $|A| = \pm 1$ or $|A||A^T| = |A|^2 = 1$, where $|A|$ denotes the determinant of A , we obtain the important group of orthogonal transformations.

The Euclidean distance between two vectors \mathbf{x}, \mathbf{y} is given by $d(\mathbf{x}, \mathbf{y}) = [(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})]^T]^{1/2}$:

$$\begin{aligned} d(\mathbf{x}', \mathbf{y}') &= [(\mathbf{x}' - \mathbf{y}') \cdot (\mathbf{x}' - \mathbf{y}')^T]^{1/2} = [(\mathbf{x} - \mathbf{y})AA^T(\mathbf{x} - \mathbf{y})^T]^{1/2} \\ &= [(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})^T]^{1/2} = d(\mathbf{x}, \mathbf{y}), \end{aligned}$$

showing that distance is also invariant under orthogonal transformations. Altogether then, areas, angles, and orthogonality are also preserved. All these together with the “pure” translation

$$\mathbf{x}' = \mathbf{x} + \mathbf{c}$$

form the *rigid motions*, which are the invariants of Euclidean geometry.

To recapitulate, as pointed out by Klein, the above geometries are projective geometries with additional conditions. For example, affine geometry is obtained by selecting any line in a projective geometry and considering only the transformations leaving that line invariant. This line plays the role of the ideal line and can be safely ignored. The reader is referred to [163], [157] for background and history of N -dimensional geometry, and to [152], [18] for learning more about the various geometries. Aspects of differential geometry, starting with a short chapter on the theory of envelopes, are covered later in the study of curves and surfaces.

♣ FT-1 The Plane with Parallel Coordinates

3.1 The Fundamental Duality

A point in the x_1x_2 plane is represented in parallel coordinates (abbreviated \parallel -coords) by a line in the xy -plane (Fig. 3.1). To find out what a line “looks like” in \parallel -coords, a set of collinear points is selected (Fig. 3.2, top right), and the lines representing them intersect (top left) at a point! Remarkably, this is true in general, and for the line

$$\ell : x_2 = mx_1 + b, \quad (3.1)$$

the corresponding point is

$$\bar{\ell} : \left(\frac{d}{1-m}, \frac{b}{1-m} \right), \quad m \neq 1, \quad (3.2)$$

where d is the distance between the parallel axes.

As will be appreciated later, the interaxis distance is *directed*, so that interchanging the axes makes the sign of d negative. The point $\bar{\ell}$ is said to *represent* the line ℓ . It is evident that a point \leftrightarrow line duality¹⁵ is at work here, which must be properly considered in the *projective* \mathbb{P}^2 rather than the Euclidean \mathbb{R}^2 plane. In (3.2), as

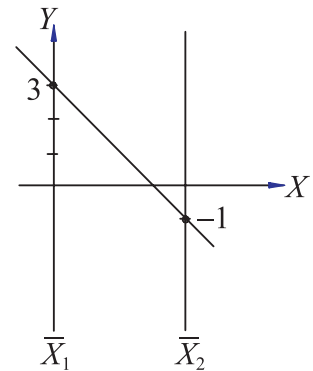


Figure 3.1. Points on the plane are represented by lines.

¹⁵See section on *duality* in Chapter 2.

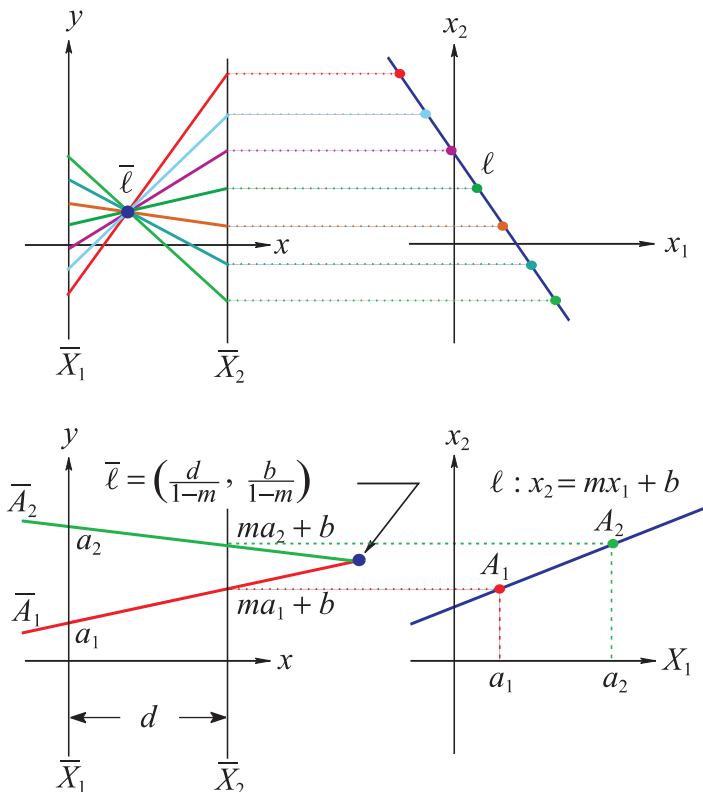


Figure 3.2. In the plane, parallel coordinates induce a point \longleftrightarrow line duality.

$m \rightarrow 1, \bar{\ell} \rightarrow \infty$ in a constant direction. This is further clarified in Fig. 3.3, showing that the lines representing points of a line with $m = 1$ are parallel with slope b/d . This is why a line with $m = 1$ is represented by the ideal point in the direction with slope b/d . Use of the projective plane allows the inclusion of lines with $m = 1$ to complete the one-to-one correspondence between lines and points $\ell \leftrightarrow \bar{\ell}$ as given by (3.1) and (3.2).

From now on, the x_1x_2 and xy planes are considered as two copies of the

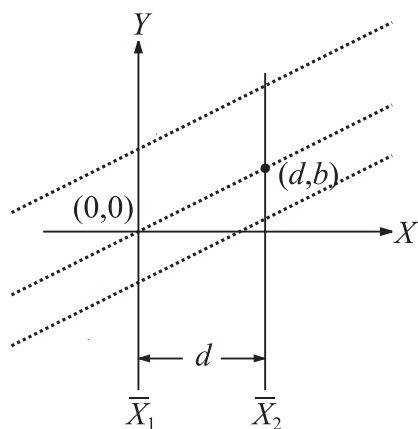


Figure 3.3. Lines representing points on the line $x_2 = x_1 + b$.

projective plane \mathbb{P}^2 . An object S in the x_1x_2 plane is represented in the xy plane by its image, denoted by \bar{S} .

Fun and Games with the ILM

Open the ILM: set *Chapter* = 2D Point Line duality.
 The display shows two sets of axes.
 On the right, click to draw a point and see the corresponding line on the left.
 Move the point and watch the line.

Next set *Subject* = *Point* \leftrightarrow *Line*.
 Draw points on the left in various x -positions and see the dual line
 on the right.

Expressing (3.2) in homogeneous coordinates with triples within $[\dots]$ and (\dots) denoting line and point coordinates respectively, we obtain

$$\ell : [m, -1, b] \longrightarrow \bar{\ell} : (d, b, 1 - m). \quad (3.3)$$

It is clear that lines with $m = 1$ are represented by the ideal points $(d, b, 0)$. The horizontal position of $\bar{\ell}$ depends only on the slope m , as illustrated in Fig. 7.11. Specifically, for vertical lines $\ell : x_1 = c, m = \infty$ and $\bar{\ell}$ is at $y = c$ on the y -axis, while for horizontal lines, $\ell : x_2 = a$ (i.e., $m = 0$), $\bar{\ell}$ is at $x = a$ (on the x -axis). Lines with negative slope are represented by points between the parallel axes, lines with $0 < m < 1$ by points to the right of the \bar{X}_2 -axis, and lines with $1 < m$ by points to the left of the \bar{X}_1 -axis. This explains why the representation of a point P is taken as the *whole* line \bar{P} , rather than a segment, for the intersection of the line \bar{P} with points $P \in \ell$ may lie outside the strip between the axes. It is clear that parallel lines having the same slope are represented by points on the same *vertical* line, a property that turns out to be very useful. Therefore, the ideal point P_m^∞ is represented by the vertical line $x = \frac{1}{1-m}$ shown in Fig. 3.5 and represents *all* the lines with slope m .

For complete generality, we use the line description

$$\ell : a_1x_1 + a_2x_2 + a_3 = 0, \quad (3.4)$$

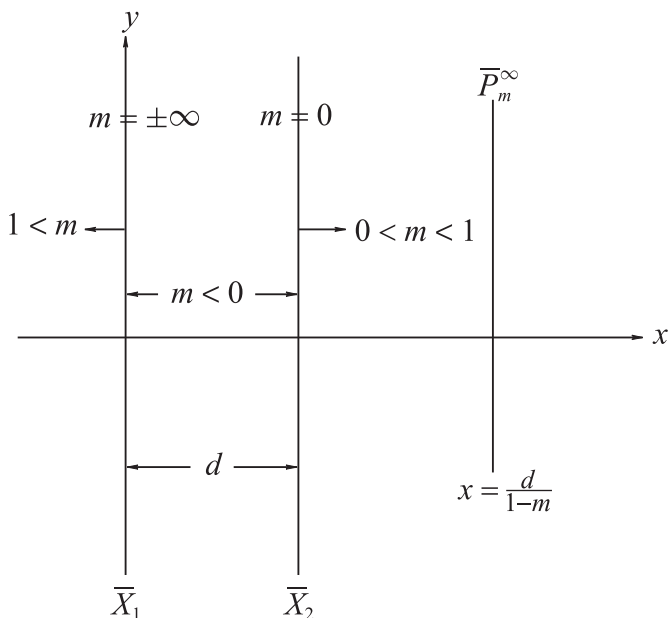


Figure 3.4. The horizontal position of $\bar{\ell}$ depends only on the slope m of ℓ . See (3.4) and (3.5) for the case $m = \pm\infty$.

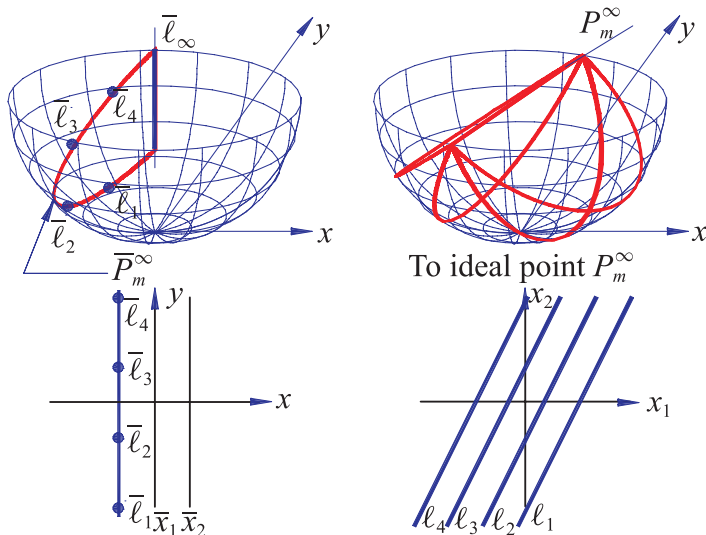


Figure 3.5. Under the duality, parallel lines map into points on the same vertical line. In the projective plane model, the great semicircles representing the lines share the same diameter, since the lines have the same ideal point.

and for $a_2 \neq 0$, we have $m = -\frac{a_1}{a_2}$, and $b = -\frac{a_3}{a_2}$, providing the correspondence

$$\ell : [a_1, a_2, a_3] \longrightarrow \bar{\ell} : (da_2, -a_3, a_1 + a_2). \quad (3.5)$$

In turn, this specifies a linear transformation between the triples ℓ and $\bar{\ell}$:

$$\bar{\ell} = A\ell, \quad \ell = A^{-1}\bar{\ell},$$

where ℓ and $\bar{\ell}$ are considered as column vectors. The 3×3 matrix is

$$A = \begin{bmatrix} 0 & d & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} -1/d & 0 & 1 \\ 1/d & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \quad (3.6)$$

which can be easily computed by taking three simple triples, such as $[1, 0, 0]$, $[0, 1, 0]$, and $[0, 0, 1]$ for ℓ .

Since $\ell \rightarrow \bar{\ell}$ must be considered in the projective plane, we can verify that the ideal line $l_\infty = [0, 0, 1]$ is mapped into $\bar{\ell}_\infty = (0, 1, 0)$, the ideal point in the vertical direction, this being the “intersection” of all the lines representing the ideal points, which we have already seen are vertical. It is clear, then, that all lines of the projective plane \mathbb{P}^2 are mapped into points of \mathbb{P}^2 as given by (3.6). Below we summarize the various cases:

$$\left\{ \begin{array}{l} \ell \rightarrow \bar{\ell} \\ [a_1, 0, a_2], m = \infty \rightarrow (0, -a_2, a_1), \text{ on the } y\text{-axis} \\ [0, a_2, a_3], m = 0 \rightarrow (d, -a_3/a_2, 1), \text{ on the } x\text{-axis} \\ [-a_2, a_2, a_3], m = 1 \rightarrow (1, -\frac{a_3}{da_2}, 0), \text{ ideal point slope } -\frac{a_3}{da_2}(\frac{m}{d}) \\ [0, 0, 1], \ell_\infty \rightarrow (0, 1, 0), \bar{\ell}_\infty \text{ vertical direction} \end{array} \right\}.$$

For the other half of the duality, using homogeneous coordinates we look carefully into the point–line $P \rightarrow \bar{P}$ correspondence illustrated in Fig. 3.6. The point $P = (p_1, p_2, p_3) = (p_1/p_3, p_2/p_3, 1)$, $p_3 \neq 0$, is represented in the xy plane by the line \bar{P} with slope $(p_2 - p_1)/dp_3$. Hence

$$\bar{P} : (dp_3)y = (p_2 - p_1)x + d p_1,$$

so that

$$P = (p_1, p_2, p_3) \longrightarrow \bar{P} = [(p_1 - p_2), dp_3, -dp_1]. \quad (3.7)$$

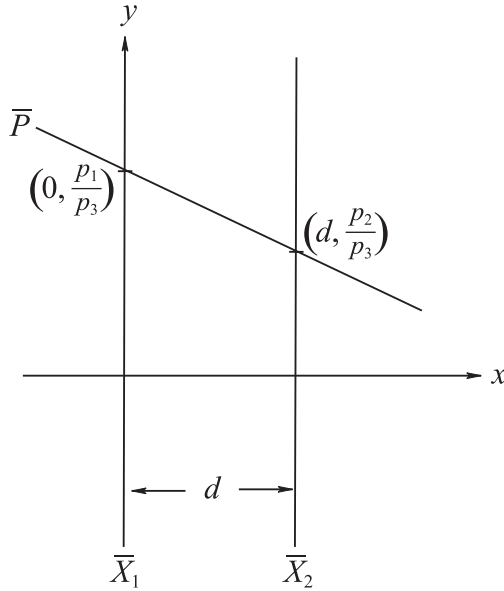


Figure 3.6. A point $P = (p_1, p_2, p_3) = (p_1/p_3, p_2/p_3, 1)$, $p_3 \neq 0$, is represented by a line \bar{P} .

The ideal point $P_m^\infty = (p_1, p_2, 0) = (1, p_2/p_1, 0)$ has direction with slope $m = p_2/p_1$. Therefore, $\bar{P}_m^\infty = [(p_1 - p_2), 0, -dp_1] = [1 - m, 0, -d]$, which is the vertical line at $x = d/(1 - m)$, as expected (see also Fig. 7.11). Considering P and \bar{P} as column vectors, we can write the above correspondence as a linear transformation:

$$\bar{P} = BP, \quad P = B^{-1}\bar{P},$$

with

$$B = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & -d \\ d & 0 & 0 \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 0 & 0 & 1/d \\ 1 & 0 & 1/d \\ 0 & -1/d & 0 \end{bmatrix}. \quad (3.8)$$

So, in \parallel -coords there is a fundamental *point* \leftrightarrow *line* duality, formalized by (3.5) and (3.7). Curiously, for $d = 1$, $A^T = B^{-1}$. Altogether, then, we have the transformations *line* $\ell \xrightarrow{A} \bar{\ell}$ *point* and *point* $P \xrightarrow{B} \bar{P}$ *line*. We can continue with $\ell \xrightarrow{A} \bar{\ell} \xrightarrow{B} \bar{\bar{\ell}}$, but in general, the line ℓ is not the same as the line $\bar{\bar{\ell}}$. That is, $B(A\ell) \neq \ell$, and likewise, $A(BP) \neq P$ (see Exercises 2, 3). By the way, the matrices A, B are quite sparse (i.e., have many zeros), and since usually $d = 1$, the computation to and from \parallel -coords is not too costly.

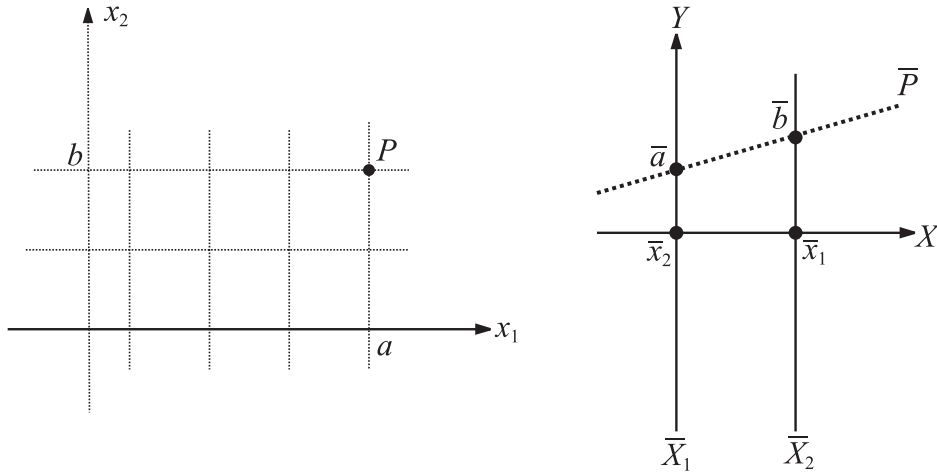


Figure 3.7. Motivation for the notation \bar{X}_i of the axes; note the position of the \bar{x}_i . Two lines $x_1 = a, x_2 = b$ on the point P (left) whose images are two points \bar{a}, \bar{b} on the line \bar{P} (right). The axis \bar{X}_1 represents the ideal point in the vertical direction (hence \bar{x}_2 is on it) and \bar{X}_2 in the horizontal (with \bar{x}_1 on it).

A few words about the notation used in Fig. 3.7 are in order. Each of the axes $\bar{X}_i, i = 1, 2$, represents a *family* of parallel lines. Specifically, \bar{X}_i represents the parallel lines $x_i = \text{constant}$ or the ideal point in the direction of x_i . Note the representation of x_1 and x_2 above.

The image (representation) \bar{r} of a curve r can be obtained as the *envelope*¹⁶ of the lines representing the points of r . This is distinguished by referring to the original curves as *point-curves* and their images as *line-curves* as in the example in Fig. 3.8. Line-curves are constructed from the *envelope* of the family of curves (for an old but very nice treatment of envelopes, see [17]). In fact, the point $\bar{\ell}$ can also be found in this way to be the envelope of the infinite family of lines \bar{A} . In Chapter 7 on curves it is seen that the mapping

$$\text{point-curve} \leftrightarrow \text{point-curve}$$

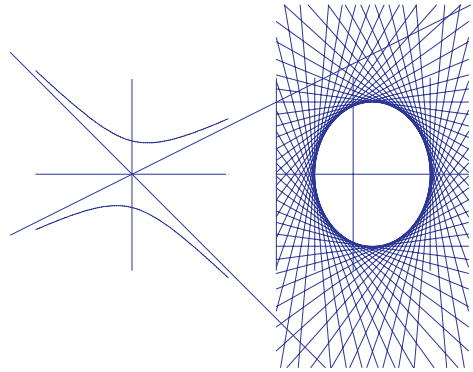


Figure 3.8. Hyperbola (point-curve) \rightarrow Ellipse (line-curve).

¹⁶As explained in the chapter on envelopes.

is achieved directly by a transformation between the x_1x_2 and xy projective planes [99]. Additional related topics can be found in [97].

Exercises

1. Verify (3.2). Hint: the lines $\bar{A}_i, i = 1, 2$, are on the points $(0, a_i), (d, ma_i + b)$ of the xy plane and $\bar{\ell} = \bar{A}_1 \cap \bar{A}_2$.
2. Apply $\ell \xrightarrow{A} \bar{\ell} \xrightarrow{B} \bar{\bar{\ell}} \rightarrow \dots$. Does the pattern repeat? That is, do we have $\ell \rightarrow \dots \rightarrow \ell$ after a finite number of transformations? What is meant here is the application of the line-to-point transformation expressed by the matrix A , then the point-to-line transformation expressed by the matrix B , etc.
3. Repeat for $P \xrightarrow{B} \bar{P} \xrightarrow{A} \bar{\bar{P}} \rightarrow \dots$. Does this pattern repeat?

Fun and Games with the ILM

Open the ILM: set *Chapter = 2D Point Line duality*,
Subject = Point \leftrightarrow Line, *Example = Horizontal Slope*,
 check *Lock Horizontal*.

On the left, move a point horizontally, while watching the rotating dual line on the right.

Change to *Example = Vertical Slope*, check *Lock Vertical*.
 Next set *Subject = Point \leftrightarrow Line*.

On the left, translate a point vertically while watching the horizontal translation of the dual line.

Change to *Example = Parallel Projective*, check *Lock Vertical*.
 There are parallel lines in Cartesian and \parallel coordinates
 and on the Projective Plane Model (PPM).

On the lower left translate a point vertically while watching the action in the other windows.

3.2 Transformations under the Duality

3.2.1 Rotations and Translations

With reference to Fig. 3.9, a line ℓ originally horizontal, shown in position ℓ_0 , is rotated counterclockwise about one of its points, which for convenience is taken as the origin O . Under the duality, the corresponding path of $\bar{\ell}$ is along the line \bar{O} , since in any of its rotated positions (say ℓ_1, ℓ_2, ℓ_3 etc.), ℓ still contains the point O . To better understand the corresponding motion of the point $\bar{\ell}$, recall that in the projective plane a line behaves like a “closed curve” (see Fig. 3.5) “joined” at its ideal point. Hence due to the counterclockwise rotation, the point $\bar{\ell}$ moves

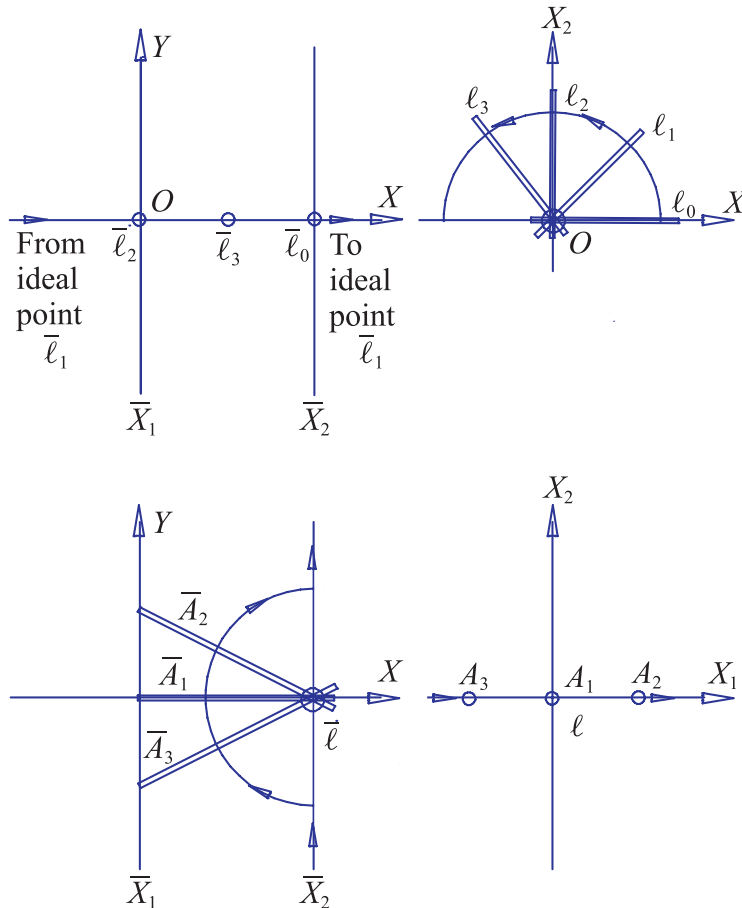


Figure 3.9. Duality between rotations and translations.

in the direction of increasing x (i.e., to the right) until the line's slope is 1 when $\bar{\ell}$ passes through the ideal point and “returns” to the Euclidean plane now, to the left of the \bar{X}_1 -axis but still moving to the right. Dually, a translation of point A in the positive direction along a line ℓ (taken for convenience coincident with the x_1 -axis) corresponds to the rotation of the line \bar{A} in the clockwise direction. The corresponding dualities in higher dimensions, which we will meet in due course, are especially useful.

Fun and Games with the ILM

Open the ILM: set *Chapter = 2D Point Line duality*,
Subject = Point \leftrightarrow Line, Example = Translation \leftrightarrow Rotation,
 check *Lock Horizontal*.

On the left move a point horizontally while watching the rotating
 dual line on the right.

Change to *Example = Rotation \leftrightarrow Translation*, check *Lock Horizontal*.
 On the right translate a point horizontally while watching the rotation
 on the left.

3.2.2 ** Recognizing Orthogonality

Parallel lines (Fig. 3.5) can be recognized from their dual point images, and so can orthogonal lines, by means of the following two transformations:

1. *Reflectio* about the line $x = \frac{1}{2}$, denoted by $R_{\frac{1}{2}}$. In the xy plane the reflection of the vertical line \bar{P}_m^∞ (the set of points representing the lines with slope m) about the line $x = \frac{1}{2}$ is the image, as shown in Fig. 3.10, of the lines with slope the reciprocal $1/m$. That is,

$$R_{\frac{1}{2}}(\bar{P}_m^\infty) = \bar{P}_{\frac{1}{m}}^\infty.$$

2. *Circle inversion*, denoted by $C_{\frac{1}{2}}$, with respect to the circle centered at $(\frac{1}{2}, 0)$ and radius $\frac{1}{2}$. We know that the vertical line \bar{P}_m^∞ at $x = 1/(1 - m)$ represents all lines with slope m . The tangent from the point $\bar{\ell}_o = \bar{P}_m^\infty \cap x$ -axis, Fig. 3.11, touches the circle at the point $\bar{\ell}_1$ whose x -coordinate is $1/(1 + m)$. So the

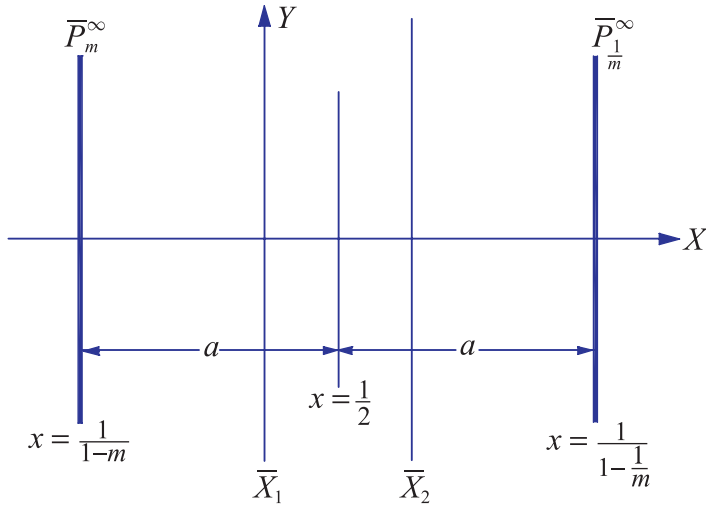


Figure 3.10. Reflection about the line $x = \frac{1}{2}$.

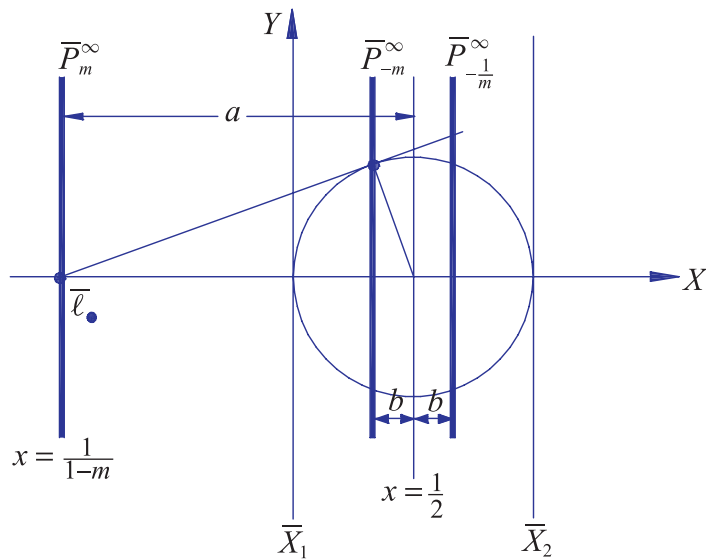


Figure 3.11. Circle inversion.

vertical¹⁷ line at $x = 1/(1+m)$ is \bar{P}_{-m}^∞ , representing all lines with slope $-m$. We write $C_{\frac{1}{2}}(\bar{P}_m^\infty) = \bar{P}_{-m}^\infty$.

¹⁷I am indebted to M. Mostrel for this result.

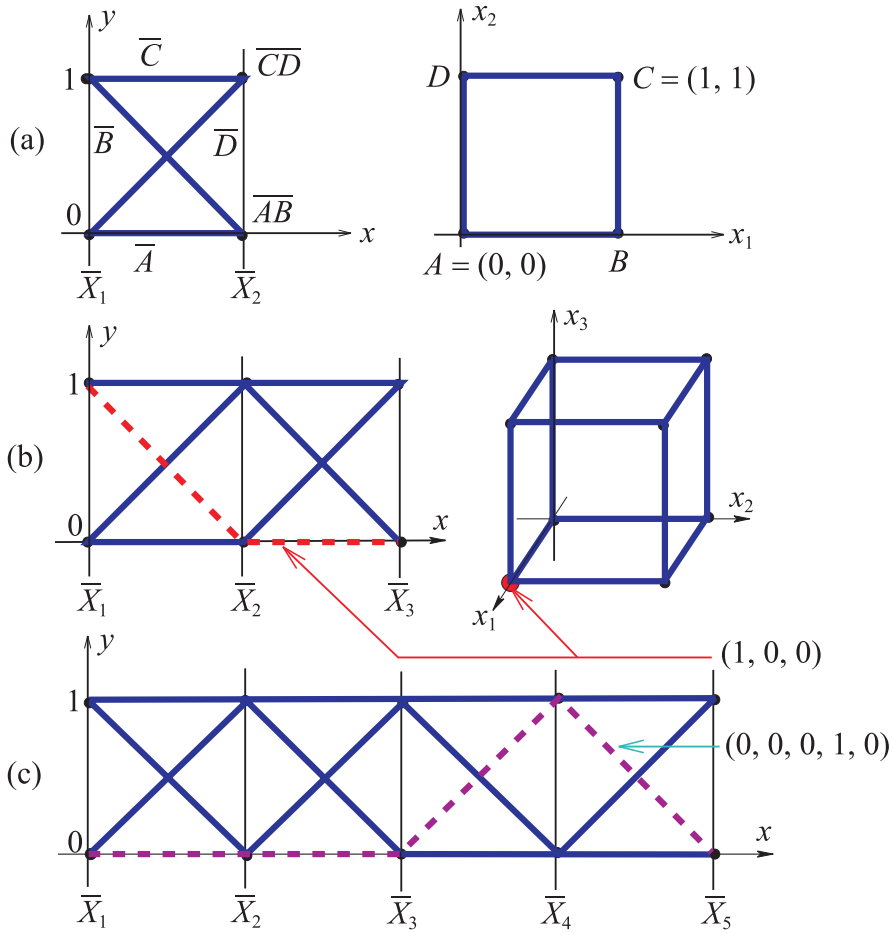


Figure 3.12. (a) Square, (b) 3-D cube, (c) 5-D hypercube, all with edges of unit length.

The family of lines *orthogonal* to those with slope m is found with the composition

$$R_{\frac{1}{2}} C_{\frac{1}{2}} (\bar{P}_m^\infty) = \bar{P}_{-\frac{1}{m}}^\infty.$$

The x -distances a, b (shown in Figures 3.10 and 3.11) of \bar{P}_m^∞ and \bar{P}_{-m}^∞ from $(\frac{1}{2}, 0)$ are

$$a = -\frac{(1-m)}{2(1+m)}, \quad b = -\frac{1+m}{2(1-m)},$$

respectively. Therefore, $ab = 1/4$ is the invariance for mutually orthogonal families of lines. All this shows that orthogonality information is preserved in the \parallel -coords representation.

3.3 A Preview

A glimpse of what lies ahead is seen in Fig. 9.1. In part (a) we see a square with unit side in Cartesian and dually in parallel coordinates. Note that the edge \overline{AB} is the point represented by the intersection of the lines \bar{A} with \bar{B} representing the corresponding two vertices. The situation becomes more interesting in part (b), where a cube with unit side and its image are shown. The collection of all vertices has coordinates that are all possible ordered triples of 0's and 1's. Therefore the cube's image consists of two adjacent copies of the square's image. Note the polygonal line indicated as the image of the vertex $E = (1, 0, 0)$, for example. It is clear that all vertices are included in the image. Also, the connection with the square and the symmetry of the cube are suggested, but not clearly understood, in the "double-square" image. In due course, we will see that this image contains *all* the information about the cube, namely, the edges and faces as well as the interior and exterior. Whereas we do not know how to show a hypercube in 5-D or higher in Cartesian coordinates, there is no problem doing so in parallel coordinates, as seen in part (c) of the figure, and with, as we shall eventually show, all the information present. We will also see that the images with repetitive patterns represent objects with high symmetry more general than the hypercube. The reason this works is that our duality has nice generalizations to higher dimensions, which is our next topic. Though the ensuing development is carried out in the projective plane, it works out that we can still use all the Euclidean notions of angle, distance, etc. as long as ideal elements are not involved.

Exercises

1. Draw in Cartesian and parallel coordinates a point $P = (p_1, p_2)$ and a line ℓ before and after:
 - (a) the translation of the origin to the point (a_1, a_2) ,
 - (b) the reflections about the x_1 and x_2 axes,
 - (c) the rotation of the axes about the origin by an angle θ ,
 - (d) the composition of a translation followed by a rotation.
2. Given a family of lines \mathcal{L} having slope m , generalize the transformations in Section 3.2.2 to construct the family of lines \mathcal{L}_θ at an angle θ with the lines in \mathcal{L} .

This page intentionally left blank

Multidimensional Lines

4.1 **FT-I** Representing Lines in \mathbb{R}^N

There are several *point* \leftrightarrow *line* dualities (such as the very useful Hough transform) in the plane that do not generalize well for higher dimensions. This is because the natural duality is *point* \leftrightarrow *hyperplane* in projective N -space \mathbb{P}^N with *point* \leftrightarrow *line* when $N = 2$. However, in \parallel -coords there is a useful and direct generalization, which is the subject of this chapter. At first the basic idea for lines in \mathbb{R}^N , rather than \mathbb{P}^N , is derived intuitively, paving the way for the general case, which is treated subsequently (as in [106] and [107]).

4.1.1 Elementary Approach

Adjacent Variables Form

What is meant by “a line in \mathbb{R}^N ”? In \mathbb{R}^3 , a line is the intersection of two planes. Generalizing, for \mathbb{R}^N , a line ℓ is the intersection of $N - 1$ nonparallel hyperplanes. Equivalently, it is the set of points (specified by N -tuples) satisfying a set of $N - 1$ linearly independent linear equations. After some manipulation, with the exception of some special cases to be treated later, the set of equations can be put in the convenient form

$$\ell \quad : \quad \left\{ \begin{array}{l} \ell_{1,2} : x_2 = m_2 x_1 + b_2, \\ \ell_{2,3} : x_3 = m_3 x_2 + b_3, \\ \dots \\ \ell_{i-1,i} : x_i = m_i x_{i-1} + b_i, \\ \dots \\ \ell_{N-1,N} : x_N = m_N x_{N-1} + b_N, \end{array} \right. \quad (4.1)$$

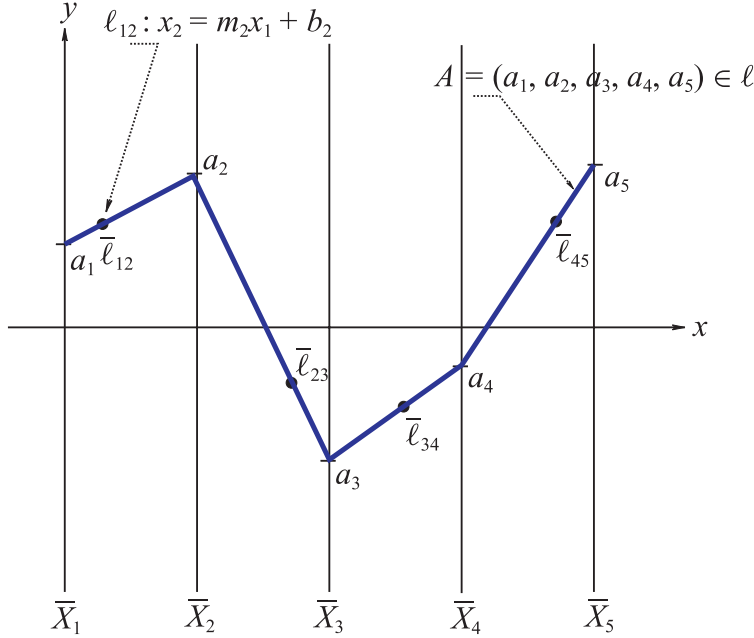


Figure 4.1. A point on a line in 5-D.

each equation containing a pair of *adjacently* indexed variables. In the $x_{i-1}x_i$ plane, the relation labeled $\ell_{i-1,i}$ is a line, and by our *point* \leftrightarrow *line* duality of Chapter 3, it can be represented by a point $\bar{\ell}_{i-1,i}$. To clarify matters, the situation is illustrated in Fig. 4.1 with an example in \mathbb{R}^5 . A straightforward extension of the 2-D duality is that a point $A = (a_1, a_2, a_3, a_4, a_5)$ is in ℓ when the polygonal line \bar{A} is on each of the points $\bar{\ell}_{12}, \bar{\ell}_{23}, \bar{\ell}_{34}, \bar{\ell}_{45}$ representing the four linear relations of (10.5) for $N = 5$. Let $\bar{A}_{(i-1),i}$ denote the $(i-1, i)$ portion (line) of \bar{A} on the points $(a_{i-1}, 0), (a_i, 0)$. Then \bar{A}_{12} on $\bar{\ell}_{12} \Rightarrow (a_1, a_2)$ satisfies the ℓ_{12} relation, and successively $(a_2, a_3), (a_3, a_4), (a_4, a_5)$ satisfy the $\ell_{23}, \ell_{34}, \ell_{45}$ relations respectively. Hence in $\|\text{-coords}$ we can *see* a line $\ell \subset \mathbb{R}^N$, as represented by $N-1$ points $\bar{\ell}$ with two indices, and its points $A \in \ell$ represented by the \bar{A} on all the $\bar{\ell}$.

Unless otherwise stated, from now on, the distance between each pair of adjacent axes is taken as one unit, as shown in Fig. 4.2. Since the y-axis is not coincident with the \bar{X}_{i-1} -axis, the x-coordinate of the point representing the line $\ell_{i-1,i}$ (see (3.3) and Fig. 3.2) needs to be translated by $(i-2)$. That is,

$$\bar{\ell}_{i-1,i} = \left(\frac{1}{(1-m_i)} + (i-2), \frac{b_i}{(1-m_i)} \right),$$

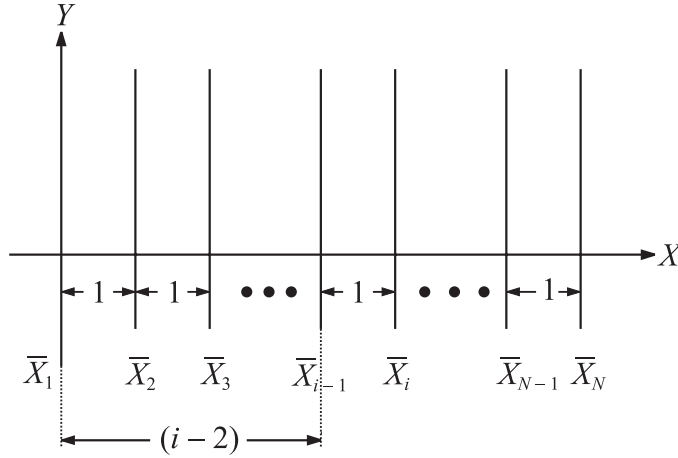


Figure 4.2. Spacing between adjacent axes is 1 unit.

and in homogeneous coordinates:

$$\bar{\ell}_{i-1,i} = ((i-2)(1-m_i) + 1, b_i, 1-m_i). \quad (4.2)$$

There are $N-1$ such points for $i = 2, \dots, N$ representing a line ℓ . This is the first instance in which the indexing arises in the representation and plays a key role in the subsequent development. For now it suffices to observe that without the indexing, the representation of the line is ambiguous. The exact number of lines that can be represented by the same points indexed differently and other related questions are studied later, in Section 4.1.4. Since the display space is at a premium, the indexing is usually not included in the picture, but it must always be accessible.

In Fig. 4.3 are several polygonal lines representing points on an interval of a line $\ell \subset \mathbb{R}^{10}$. The intersection of the polygonal lines between the \bar{X}_2'' and \bar{X}_3 -axes is the point $\bar{\ell}_{2,3}$. All nine points representing that 10-dimensional line can be seen (or constructed) with their horizontal positions depending on the first coordinate of (4.2). Do not be misled by the fact that here, all of the $\bar{\ell}$'s except $\bar{\ell}_{1,2}$ and $\bar{\ell}_{6,7}$ are between their corresponding axes. This is due to the choice of $m_i \leq 0$ made for display convenience only. As for 2-D, the x -coordinate of $\bar{\ell}_{i-1,i}$ is between the \bar{X}_{i-1} , \bar{X}_i axes when $m_i < 0$, to the left of \bar{X}_{i-1} when $m_i > 1$, and to the right of \bar{X}_i when $0 < m_i < 1$.

Base Variable Form

Another common way of describing a line $\ell \subset \mathbb{R}^N$ is in terms of one variable, sometimes called the *base*, which after appropriate relabeling may be taken as x_1 .

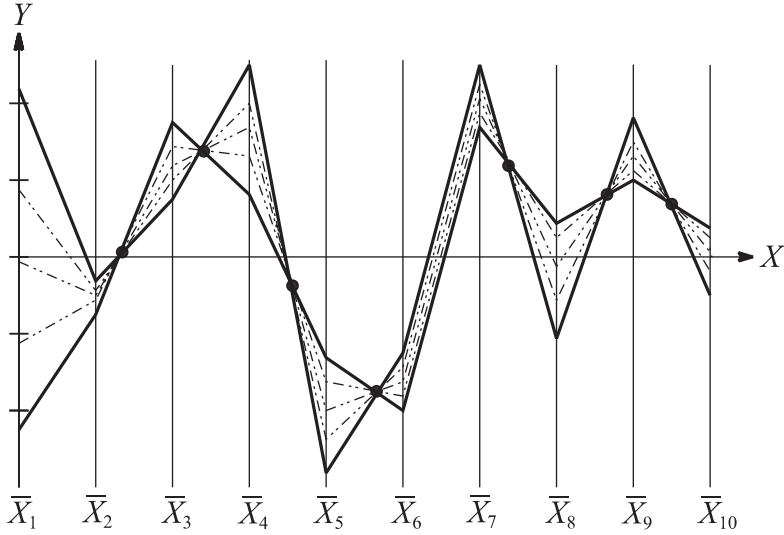


Figure 4.3. Line interval in \mathbb{R}^{10} .

The thicker polygonal lines represent its endpoints. The adjacent-variables representation, consisting of nine properly indexed points, is obtained by the sequential intersections of the polygonal lines. Note that $\ell_{1,2}$ is to the right of the X_2 -axis, and $\bar{\ell}_{6,7}$ is an ideal point. The remaining points are between the corresponding pairs of axes.

Then

$$\ell : \begin{cases} \ell_{1,2} : x_2 = m_2^1 x_1 + b_2^1, \\ \ell_{1,3} : x_3 = m_3^1 x_1 + b_3^1, \\ \dots \\ \ell_{1,i} : x_i = m_i^1 x_1 + b_i^1, \\ \dots \\ \ell_{1,N} : x_N = m_N^1 x_1 + b_N^1, \end{cases} \quad (4.3)$$

and the $N - 1$ points representing the line are

$$\bar{\ell}_{1,i} = (i - 1, b_i^1, 1 - m_i^1). \quad (4.4)$$

With the axes placed 1 unit apart, the distance between \bar{X}_i and \bar{X}_1 (which is coincident with the y -axis) is $i - 1$. This brings the point $\bar{\ell}_{1,i}$ between the axes \bar{X}_1, \bar{X}_i when $m_i^1 < 0$, to the left of \bar{X}_1 when $m_i^1 > 1$, and to the right of \bar{X}_i when $0 < m_i^1 < 1$. All the remarks about indexing pertain equally well to this or any other representation of ℓ .

What about changing the parametrization? Let us consider an example in which a line in $\ell \subset \mathbb{R}^5$ is specified by four indexed points of the form $\bar{\ell}_{i-1,i}$ and it is required to find the linear relation between x_2 and x_5 :

$$\ell_{2,5} : x_5 = m_{25}x_2 + b_{25}. \quad (4.5)$$

There is no problem returning to the system of four linear equations $\ell_{i-1,i}$, $i = 2, \dots, 5$, specifying the line ℓ , from which the points $\bar{\ell}_{i-1,i}$ were obtained, doing the necessary algebraic manipulations to obtain (4.5). But this can be done equivalently by the geometrical construction shown in Fig. 4.4. By assigning any two values to x_1 , the polygonal lines representing two points on ℓ are constructed. Joining the two pairs of x_2 and x_5 values provides two lines whose intersection must, due to the *line* \rightarrow *point* mapping in 2-D, be the point $\bar{\ell}_{2,5}$ with coordinates

$$\bar{\ell}_{25} = \left(\frac{3}{1 - m_{25}} + 1, \frac{b_{25}}{1 - m_{25}} \right), \quad (4.6)$$

providing the values of m_{25} and b_{25} . By the way, the 3 is for the distance between the \bar{X}_2 - and \bar{X}_5 -axes, and the translation by 1 is for the distance between the \bar{X}_2 - and y -axes. Incidentally, we see that all the information about a line is really contained in its $N - 1$ indexed points. For the remainder, all sections marked by ** can be safely omitted at first reading.

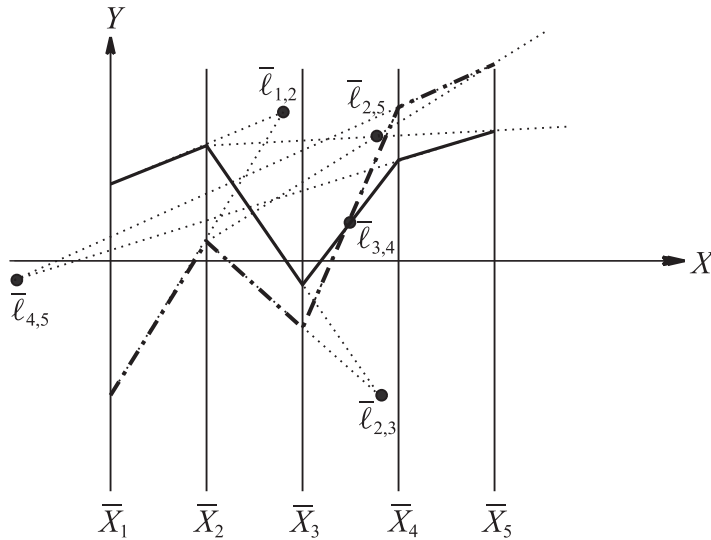


Figure 4.4. Algorithm for constructing a pairwise linear relation. Here $\bar{\ell}_{25}$ is constructed from the four points $\bar{\ell}_{i-1,i}$ representing the line.

Fun and Games with the ILM

set *Chapter = Multidimensional Lines*, *Subject = Multidimensional*,
Example = Indexed Points.

The two points (polygonal lines) in 6-D specify a line.
 The *adjacent variables* indexed points for the line are shown.

For some experimentation:
 change any of the points' coordinates by moving it along its axis and
 observe the change in **two** of the indexed points. Repeat with other variables.
 Try to make two, three or more of the indexed points coincident.

Next click *adjacent variables* off and *base variable* on.
 Repeat the above experimentation for the *base variable* indexed points.
 In anticipation of the result in Section 4.1.2,
 experiment with the *collinearity* option by changing the index triples.



Two-Point Form

The representation of a line ℓ independent of parametrization can best be constructed from two of its points. Starting with

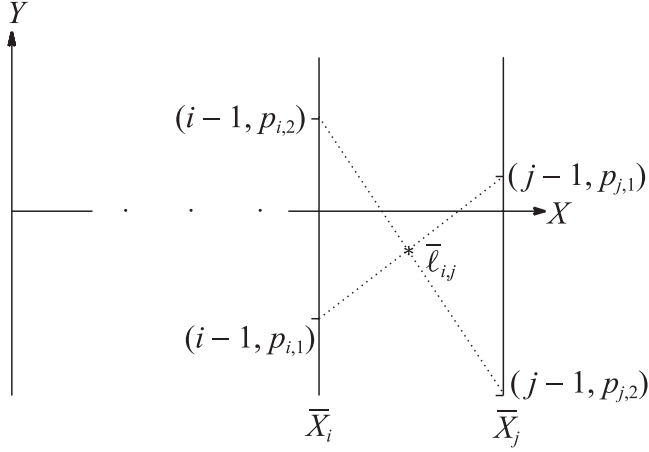
$$P_r : (p_{1,r}, p_{2,r}, \dots, p_{N,r}), \quad r = 1, 2,$$

for any other point $P \in \ell$, we have $P = \lambda P_1 + (1 - \lambda)P_2, \lambda \in \mathbb{R}$. When both points are ideal, they define an ideal line. Otherwise, if one of the points is ideal, it is a simple matter to replace it by a second ordinary point distinct from the first to obtain the same line ℓ . So without loss of generality, only lines defined by two ordinary points need to be dealt with.

Two pairs

$$(p_{i,r}, p_{j,r}), \quad i, j \in 1, 2, \dots, N, \quad r \in 1, 2, \quad (4.7)$$

specify the projection ℓ_{ij} of ℓ on the $x_i x_j$ coordinate plane, which is represented by a point $\bar{\ell}_{i,j}$. It is essential to identify it with the proper indices as in Fig. 4.5.

**Figure 4.5.** Construction of $\bar{\ell}_{i,j}$.

For the data given in (4.7), we have

$$\ell_{ij} : \Delta_j x_i - \Delta_i x_j + D_{i,j} = 0, \quad (4.8)$$

with line coordinates $[\Delta_j, -\Delta_i, D_{i,j}]$ for

$$\begin{cases} \Delta_i = p_{i,2} - p_{i,1}, \\ \Delta_j = p_{j,2} - p_{j,1}, \\ D_{i,j} = p_{j,2}\Delta_i - p_{i,2}\Delta_j = p_{j,1}\Delta_i - p_{i,1}\Delta_j, \end{cases} \quad (4.9)$$

for all $i, j \leq N$. When Δ_i and Δ_j are not both zero, (4.8) actually defines a line that is represented by the point

$$\bar{\ell}_{i,j} : ((i-1)\Delta_j - (j-1)\Delta_i, -D_{i,j}, \Delta_j - \Delta_i), \quad (4.10)$$

where as usual, the distance between adjacent coordinate axes is 1. The relevant transformations here are

$$\bar{\ell}_{i,j} = A_{i,j} \ell_{i,j}, \quad \ell_{i,j} = A_{i,j}^{-1} \bar{\ell}_{i,j}, \quad (4.11)$$

where

$$A_{ij} = \begin{pmatrix} (i-1) & (j-1) & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

and

$$(i - j)A_{ij}^{-1} = \begin{pmatrix} 1 & 0 & (1 - j) \\ -1 & 0 & (i - 1) \\ 0 & (j - 1) & 0 \end{pmatrix}.$$

There are still some important considerations missing, which are completed in Section 4.1.4.

4.1.2 **FT-2** Some Properties of the Indexed Points

The 3-point Collinearity

The representation of a line in terms of $N - 1$ points with two indices can be given for *any* of the various line descriptions by linear equations each involving a pair of variables. Once a sufficient representation in terms of $N - 1$ points is given, any other point $\bar{\ell}_{i,j}$ can be obtained from the representation by a geometric construction as in Fig. 4.4. It is important to clarify that by the *representation* \bar{S} of an object S is meant the *minimal* subset of the xy plane from which the representative \bar{P} of any point $P \in S$ can be constructed. For a line $\ell \subset \mathbb{P}^N$ we have seen that $N(N - 1)$ indexed points in the xy plane, corresponding to all distinct pairs of the N variables, can be constructed. However only $N - 1$ of these, corresponding to $N - 1$ linearly independent equations in pairs of these variables, are needed to *represent* the line ℓ .

The indexed points representing the linear relation between pairs of variables for an N -dimensional line have a striking and very useful property. For $i \neq j \neq k \in [1, 2, \dots, N]$, the three points $\bar{\ell}_{i,j}, \bar{\ell}_{j,k}, \bar{\ell}_{i,k}$ are always collinear. This can be seen by considering again two points $P_r = (p_{1,r}, \dots, p_{N,r}), r = 1, 2$, on ℓ and their projections on the x_i, x_j, x_k three-space as shown in Fig. 4.6. The projected portions of the points are the vertices of two triangles, with the collinearity of $\bar{\ell}_{i,j}, \bar{\ell}_{j,k}, \bar{\ell}_{i,k}$ being a consequence of Desargues's theorem.¹⁸ Here the two triangles are in perspective with respect to the ideal point in the vertical direction. This property, as will be seen, is the “backbone” of the construction for the representation of higher-dimensional p -flats in N -space (that is, planes of dimension $2 \leq p \leq N - 1$), and it is referred to as the *the 3-point collinearity property*.

¹⁸Two triangles in perspective from a point are in perspective from a line.

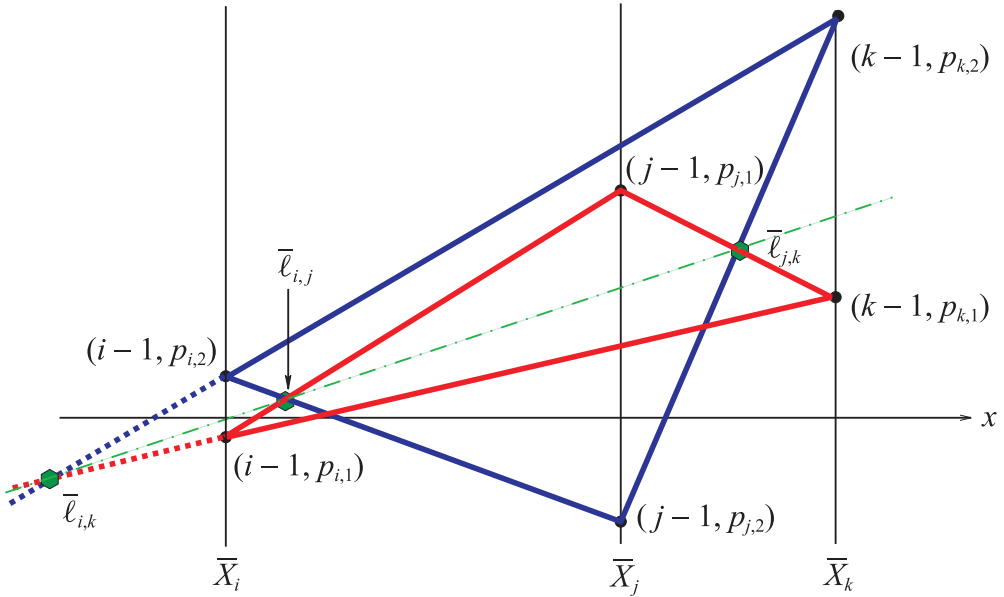


Figure 4.6. The collinearity for the three points $\bar{\ell}_{i,j}$, $\bar{\ell}_{j,k}$, $\bar{\ell}_{i,k}$. The two indicated triangles (blue and red) are in perspective with respect to the ideal point in the vertical direction. The y-axis is off scale.

Fun and Games with the ILM

set Chapter = Multidimensional Lines, Subject = Multidimensional,
Example = Collinearity.

The two points (polygonal lines) in 6-D specifying a line appear.
Three indexed points are collinear.

Change any point's coordinates by moving it along its axis.
Observe that the collinearity is preserved.

Choose other triples of variable indices and follow the collinearity property.

Lines in a Plane

An important special case is that for any line ℓ in \mathbb{R}^3 , the three points $\bar{\ell}_{1,2}$, $\bar{\ell}_{1,3}$, $\bar{\ell}_{2,3}$ are always collinear. In turn, this provides an important corollary.

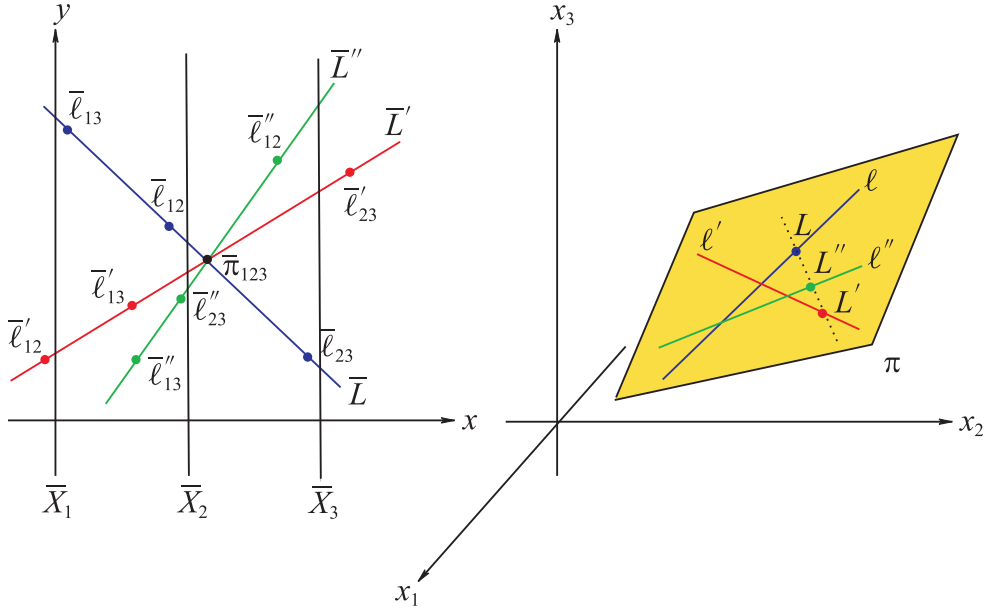


Figure 4.7. Three lines ℓ, ℓ', ℓ'' in a plane $\pi \subset \mathbb{R}^3$.

In \parallel -coords, the three-point collinearity property applied to the three points $\bar{\ell}_{12}, \bar{\ell}_{13}, \bar{\ell}_{23}$ determines a line \bar{L} , and similarly, ℓ' determines the line \bar{L}' . Denoting the intersection point by $\bar{\pi}_{123} = \bar{L} \cap \bar{L}'$, the line \bar{L}'' , determined by *any* other line ℓ'' , is *on* the point $\bar{\pi}_{123}$. The picture hints that the points L, L', L'' are on a line in the plane π .

Consider the plane

$$\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0,$$

and choose two lines $\ell, \ell' \subset \pi$, denoting by \bar{L}, \bar{L}' the lines determined by $\bar{\ell}_{1,2}, \bar{\ell}_{2,3}, \bar{\ell}_{1,3}$ and $\bar{\ell}'_{1,2}, \bar{\ell}'_{2,3}, \bar{\ell}'_{1,3}$ respectively; see Fig. 4.7. The coordinates of $\bar{\pi}_{123} = \bar{L} \cap \bar{L}'$ turn out to be

$$\bar{\pi}_{123} = (c_2 + 2c_3, c_0, c_1 + c_2 + c_3). \quad (4.12)$$

They depend on the coefficients of π and not on the choice of ℓ and ℓ' . Hence, for any other line $\ell' \subset \pi$, as in Fig. 4.7, the corresponding \bar{L}'' must also be on $\bar{\pi}_{123}$. This is a preview of the next chapter, where it is proved that two such points each with *three indices* represent a plane π in \mathbb{R}^3 .

Exercises

1. Consider a line $\ell \subset \mathbb{R}^N$ whose description, by a set of linear equations, contains $x_i = c_i$, a constant, for some i .
 - (a) How would such a line be represented?
 - (b) Can such a line be represented by $N - 1$ indexed points?
2. Prove the property illustrated in Fig. 4.7
3. Given a line ℓ and the $N - 1$ points $\ell_{i-1,i}$, $i = 2, \dots, N$, provide an algorithm for determining the point ℓ_{ij} , for any distinct pair i, j , and state its complexity; see Fig. 4.4.
4. For lines such that some of the m 's are close to 1, their representation becomes problematic. Find a transformation that can always convert points outside the axes to equivalent points between the axes. Write an algorithm to perform this transformation so that all points representing a line are between the axes. What is the algorithm's complexity?

♣ FT-2e

4.1.3 Representation Mapping I

The 3-point collinearity property and its higher-dimensional relatives are utilized in the representation of p -flats, planes of dimension $0 \leq p \leq N - 1$, in \mathbb{R}^N by means of indexed points. The 0 lower bound pertains to points $P \in \mathbb{P}^N$.

The method's goal is the construction of a mapping \mathcal{J} for the *unique representation* of objects (i.e., subsets) of \mathbb{P}^N by subsets of \mathbb{P}^2 . The object $B \subset \mathbb{P}^N$ is *represented* by its image $\mathcal{J}(B) = \bar{B}$. Starting with the representation of a point by a polygonal line $P \in \mathbb{P}^N$, it is clear that \mathcal{J} is *not a point-to-point mapping*. Formally,

$$\mathcal{J} : 2^{P^N} \rightarrow 2^{P^2} \times 2^{[1,2,\dots,N]}, \quad (4.13)$$

where $2^A = [B \subseteq A]$ denotes the *power set* of a set A , the set of all subsets of A , and \mathbb{R}^N is embedded in the projective N -space \mathbb{P}^N . The $2^{[1,2,\dots,N]}$ in the product pertains to the indexing by subsets of $[1, 2, \dots, N]$.

The answer to “how can \mathbb{R}^N be mapped into the plane \mathbb{R}^2 without loss of information?” is found by examining the *cardinality*¹⁹ of the domain and range

¹⁹Only elementary considerations are needed. For example, the sets of rational numbers and real numbers \mathbb{R} have cardinalities \aleph_0 and \aleph_1 respectively. The comparison $\aleph_0 < \aleph_1$ signifies the *nonexistence* of a one-to-one and onto mapping from the rationals to the reals. For more on the fascinating topic of infinite sets and cardinal numbers see any good text on set theory.

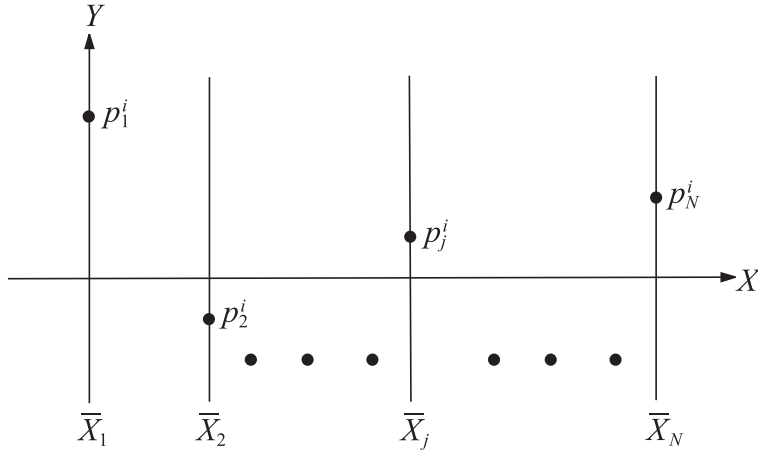


Figure 4.8. A point $P^i = (p_1^i, p_2^i, \dots, p_j^i, \dots, p_N^i)$ represented by N points p_j^i with one index.

of \mathcal{J} . Let $|A|$ denote the cardinality of a set A and recall that by taking the finite product of a set with itself the cardinality remains invariant. For our specific case,

$$|\mathbb{R}| = |\mathbb{R}^2| = |\mathbb{R}^N| = \aleph_1 \quad \Rightarrow \quad |2^{\mathbb{R}^2}| = |2^{\mathbb{R}^N}| = \aleph_2.$$

With the domain and range having the same cardinality, the construction of a one-to-one mapping \mathcal{J} is possible; what is left is to actually do it. “The devil is in the details,” as sages have told. The image $\mathcal{J}(B)$ of a set $B \subset \mathbb{P}^N$ is given in terms of the xy Cartesian coordinates superimposed on the parallel axes with $\mathcal{J}(B) = \mathcal{J}(C) \Leftrightarrow B = C$. At this stage we can get a glimpse of the mapping’s *recursive construction algorithm*, the recursion being on the dimensionality of the object being represented.

1. *The nonrecursive step* is the alternate representation of a point

$$P^i = (p_1^i, p_2^i, \dots, p_j^i, \dots, p_N^i).$$

Rather than using a polygonal line for \bar{P}^i , for consistency with what follows, the representation is also given in terms of indexed points in the xy plane. Specifically, $P^i \in \mathbb{P}^N$ is mapped into (or represented by) N points with *one index*, as shown in Fig. 4.8. Of course, this is completely equivalent to its polygonal line representation, where the indexing sequence provides the straight-line connectivity when needed.

If N_r denotes the number of points and n_r the number of indices appearing in the representation, for a (point) 0-*flat* ,

$$N_r + n_r = N + 1.$$

2. The *first recursive step* is the construction of a 1-*flat* (line) $\ell \subset \mathbb{P}^N$ consisting of $N - 1$ points with two indices. The steps are:
 - (a) for each of two distinct 0-*flat* $P_1, P_2 \subset \ell$ the N representing points are connected via the indexing sequence to provide two polygonal lines,
 - (b) intersecting adjacent portions of the polygonal lines yields the $N - 1$ points $\bar{\ell}_{i-1,i}, i = 2, \dots, N$.

Alternative but equivalent representations correspond to the intersection of $N - 1$ pairs of lines joining the appropriate alternate vertices of the polygonal lines.

Checking the previous sum for 1-*flat* yields

$$N_r + n_r = (N - 1) + 2 = N + 1.$$

3. The construction procedure turns out to be generic for the representation of p -flats of all dimensions.

The emerging relation for $N_r + n_r$ holds with the caveat that the subset of points in the representation is minimal. It is revisited as new representations are obtained.

4.1.4 ** The General Case

Indexed Points for Arbitrary Parametrization

Though the basics have been clarified, there remain representational issues due to the orientation of some lines with respect to the principal axes. Specifically, a line ℓ may be perpendicular to a principal 2-plane $x_i x_j$. Then both x_i and x_j are constant, and the projection of $\ell_{i,j}$ on the $x_i x_j$ plane is a point (p_i, p_j) represented in parallel coordinates by a line on the two points $(i - 1, p_i), (j - 1, p_j)$, which plays the role of $\bar{\ell}_{i,j}$. These contingencies, though conceptually straightforward, require that the foundations be revisited in a thorough way.

In \mathbb{R}^N there are $N(N - 1)$ different linear equations connecting the variables pairwise, and therefore, as N increases, the number of sets of equations available for describing a line increases rapidly. A precise count as a function of N is given shortly. The reasons for choosing one description rather than another are irrelevant here. Rather, the *class* of such descriptions is of interest, for which purpose a graph-theoretic approach is appropriate. Let the N variables be represented by N vertices

of a graph. An edge connecting a pair of vertices indicates that a linear relation is given between the corresponding pair of variables. A set of equations describing a line ℓ corresponds to a subgraph with N vertices and $N - 1$ edges, which must be a *spanning tree* of the complete graph of N vertices. For if the subgraph is not connected, the corresponding system of equations contains at least two independent sets of variables such that their values are independent of values taken on by the variables of other sets. Also, if there is a closed loop in the graph, then at least one vertex is disconnected, there being only $N - 1$ edges. Once a particular variable is selected and designated as the *root* of the tree, the number of different ways in which N vertices and $N - 1$ edges can be assembled into a spanning tree is $N^{(N-1)}$ (Cayley's formula) [150]. In [54] there is a nice proof²⁰ that the number of spanning trees with N distinct vertices is $N^{(N-2)}$. From this result, Cayley's formula can be immediately obtained by choosing any one of the N vertices as the root. From our viewpoint this is the number of distinct $N - 1$ linearly independent equations, with the chosen variable as the root, that can be selected out of the possible $N(N - 1)$ linear equations between pairs of variables. Assigning a value to the *start* variable, corresponding to the root, *initiates* the computation and suffices to determine the values of all the remaining variables for the specification of a point on the line.

Since variables that are constant are not involved in linear relations with another variable, there is no edge connecting their vertex to another vertex. That is to say, such vertices are isolated. If there are $N - M$ such variables, then the remaining M vertices together with $M - 1$ edges again form a spanning tree as before, *rooted* once a start variable is chosen. Of course, in this case there are $M^{(M-1)}$ ways for selecting the $M - 1$ independent linear equations, between pairs of variables, with the designated *start* variable. Let S be the set of indices for which the variables are not constant, where

$$S = (i_1, i_2, \dots, i_M)$$

and $1 \leq i_1 < i_2 < \dots < i_M \leq N$, and let \tilde{S} be the complementary set of indices. For the example shown in Fig. 4.9, $S = (1, 3, 4, 5, 8, 9)$ and $\tilde{S} = (2, 6, 7)$. Note that two lines ℓ_1 and ℓ_2 are parallel if and only if

- $S_1 \equiv S_2 = S$ and
- for all $i_r \neq i_s$ in S_1 the $\bar{\ell}_{i_r, i_s}$ have the same abscissa for both lines.

In other words, the $\bar{\ell}_{i_r, i_s}$ for each pair $i_r, i_s \in S$ lie on a vertical line as in the 2-D case discussed in Chapter 3. Clearly, consideration of $\tilde{S}_i, i = 1, 2$, is not needed, since x_i and x_j constant implies that ℓ is orthogonal to the $x_i x_j$ plane.

²⁰Also cited in [54] is [137], which has a collection of different proofs for the same result.

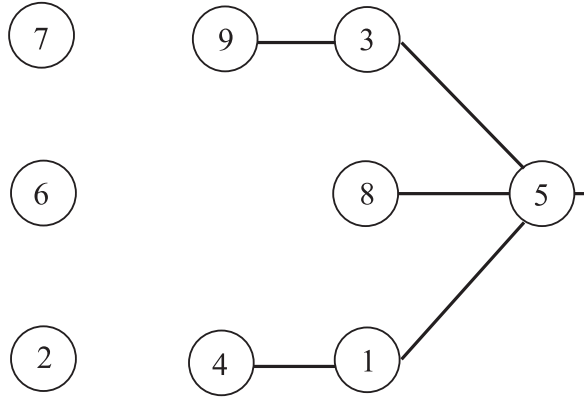


Figure 4.9. Tree with 5 as the root corresponds to x_5 being the *start* variable. The isolated vertices have indices $\tilde{S} = (2, 6, 7)$.

We have shown that any line $\ell \subset \mathbb{R}^N$ is given, after an appropriate relabeling of the variables, by the $N - 1$ equations of the form

$$\ell_{i,j} : x_j = m_i^j x_i + b_i^j, i < j \in S = (i_1, i_2, \dots, i_M), \quad (4.14)$$

$$\ell_{o,k} : x_k = b_k, k \in \tilde{S} = (i_{M+1}, \dots, i_N). \quad (4.15)$$

The “dummy” index o is introduced to preserve uniformity in indexing each relation by *two* indices rather than sometimes by two and at other times (i.e., the constant variables) by *one*. From this and the considerations provided in Section 4.1.1, the following theorem is immediate:

Theorem 4.1.1 (Representation of a line $\ell \subset \mathbb{R}^N$ for arbitrary parametrization) *Any line $\ell \subset \mathbb{R}^N$ can be represented by $N - 1$ points each of which has two indices, where $M - 1$ points are given by*

$$\begin{cases} \bar{\ell}_{i,j} = (j - i, b_i^j, 1 - m_i^j) \text{ for } i, j \in S = (i_1, i_2, \dots, i_M), i \neq j, \text{ and} \\ \bar{\ell}_{o,k} = (k - 1, b_k, 1) \text{ for } k \in \tilde{S} = (i_{M+1}, \dots, i_N). \end{cases} \quad (4.16)$$

As in the previous cases, it is clear that the horizontal position of the point $\bar{\ell}_{i,j}$ is determined by the slope m_i^j . Specifically, for $i < j$, $\bar{\ell}_{i,j}$ is to the left of the \bar{X}_i -axis for $m_i^j > 1$, between the \bar{X}_i - and \bar{X}_j -axes for $m_i^j < 0$, and to the right of the \bar{X}_j -axis for $0 < m_i^j < 1$. The construction of a point on ℓ given $\bar{\ell}$ is illustrated

in Section 4.1.5. Once a value is assigned to any variable with index in S via the $\bar{\ell}_{i,j}$'s, all other variables with indices in S are known. Since variables in \tilde{S} are also known, all the N coordinates of a point on ℓ are known.

Two-Point Form

The general *two-point form* representation, obtained in a completely analogous way, is given by the following theorem:

Theorem 4.1.2 (Representation of a line $\ell \subset \mathbb{R}^N$: Two-Point Form) *Any line in \mathbb{R}^N can be represented by $N - 1$ points, where $M - 1$ points (some possibly ideal), one for each edge connecting variable i and variable j , are given by*

$$\bar{\ell}_{i,j} = (x_{i,j}, y_{i,j}),$$

where for $\Delta_i \neq \Delta_j$ the point coordinates are

$$x_{i,j} = \frac{(j-i)\Delta_i - (i-1)(\Delta_i - \Delta_j)}{\Delta_i - \Delta_j}, \quad y_{i,j} = \frac{p_{i,1}p_{j,2} - p_{i,2}p_{j,1}}{\Delta_j - \Delta_i}, \quad (4.17)$$

and for $\Delta_i = \Delta_j$, the ideal point (direction) is

$$x_{i,j} = j - i, \quad y_{i,j} = p_{j,1} - p_{i,1}. \quad (4.18)$$

In this case, $p_{j,2} - p_{i,2} = p_{j,1} - p_{i,1}$, and there $N - M$ points given by

$$(k - 1, x_k), \quad (4.19)$$

one for each $k \in \tilde{S}$. By the definition of $\bar{\ell}_{i,j}$, the three points $(i - 1, x_i)$, $\bar{\ell}_{i,j}$, $(j - 1, x_j)$ lie on a straight line, so that as a consequence of (4.17),

$$(i - 1 - x_{i,j})x_j + (j - i)y_{i,j} - (j - 1 - x_{i,j})x_i = 0, \quad (4.20)$$

and as a consequence of (4.18),

$$x_{i,j}x_j - x_{i,j}x_i = (j - i)y_{i,j}. \quad (4.21)$$

Once a value is assigned to any variable with index in S , then via the $\bar{\ell}_{i,j}$'s all other variables with indices in S are found. But variables in \tilde{S} being constant are also known, and so are all the N coordinates of a point. Given x_i let λ satisfy the equation

$$x_i = \lambda p_{i,1} + (1 - \lambda)p_{i,2}, \quad (4.22)$$

which is always possible, since by hypothesis, $p_{i,1} \neq p_{i,2}$. It follows directly from (4.17), (4.18), (4.20), and (4.21) that

$$x_j = \lambda p_{j,1} + (1 - \lambda)p_{j,2}. \quad (4.23)$$

That is, any point so defined satisfies the definition and belongs to the straight line.

4.1.5 ** Construction Algorithms

And now, after all this preparation, we provide “pencil and paper” construction algorithms for viewing the objects that we have been discussing. This entails the representation of the line (in terms of indexed points), as well as the construction of points on the line. Together with the rooted tree representation for the desired parametrization (i.e., the $\ell_{i,j}$) and the isolated vertices corresponding to the variables with constant values, an array D with $N - 1$ rows and five columns is used as data structure. Essentially, the order of rows is unimportant, although programming considerations may dictate otherwise. There are two distinct types of rows in the structure, i.e.,

$$\begin{array}{ccccc} i & j & x_{i,j} & y_{i,j} & h_{i,j} \\ 0 & k & 0 & b_k & 1 \end{array}$$

The first type provides the indices of variables connected by an edge in the first two columns. The next three columns are the homogeneous coordinates $(x_{i,j}, y_{i,j}, h_{i,j})$ of the point $\bar{\ell}_{i,j}$, $h_{i,j} = 1$ for regular (i.e., finite) and $h_{i,j} = 0$ when it is ideal, in which case the coordinates $(x_{i,j}, y_{i,j})$ define the slope of the direction. The second row type pertains to $x_k = b_k$ a constant, with $x_{i,j} = 0$, $y_{i,j} = b_k$, and $h_{i,j} = 1$. In other words, the variable’s index is in column two, and the variable’s value is in column four. Columns three and five have only 0’s and 1’s respectively.

In the complete data structure there are $M - 1$ rows of the first type and $N - M$ rows of the second type. Of course, for fixed variables there is some waste in the structure. But since the occurrence of such variables will certainly be a rare event, it is not useful to maintain this part of the data in a separate structure. For the time being it is assumed that no Δ_i is zero, that is, all variables are unconstrained. It is apparent that there is sufficient data in the data structure to compute a point on the line. To guide the computation of such a point, a framework, called *adjacency data*, is provided. It consists of a set of arrays denoted by A_i , one for each $i \in S$. The array A_i has two rows and as many columns as there are edges connecting the i th variable to other variables having the form

$$A_i = \begin{pmatrix} i_1 & j_2 & \cdots & i_j \\ r_1 & r_2 & \cdots & r_j \end{pmatrix}.$$

The first column of the array signifies that there is an edge connecting variables with indices i and i_1 and that $\bar{\ell}_{i,i_1}$ is found in columns 3, 4, 5 of row r_1 of the data structure, and likewise for the remaining columns of the array. Note that reordering rows of the data structure requires a like reordering of the *second* rows of the adjacency arrays. The algorithm constructing a point on a line can now be stated.

Input: the adjacency data, the data structure, and an index k for which x_k , the *start variable*, is to be assigned a value v_k .

Output: a list of N values, the coordinates of a point on a line.

1. *for* $i = 1$ to N mark all x 's new
2. put k on queue (Q), *for* v_k on $x_k \leftarrow \text{old}$
3. *while* Q not empty *do*
4. *for* adjacency array A_i *for* first i in Q
 - (a) *while* adjacency array not empty *do*
 - (b) *for* x_j not old, j first in row 1
 - i. find the value of x_j
 - ii. append j to Q
 - iii. $x_j \leftarrow \text{old}$
 - iv. delete first column of A
 - (c) delete first index in Q
5. stop

The treatment of the queue Q is such that each variable is dealt with at most once. The graph is connected, and via the adjacency data, each variable enters the queue. Thus each variable is dealt with exactly once. Given the value of one variable, there are $N - 1$ calculations required for the remaining variables. Now each edge (i, j) appears twice in the adjacency data, once in A_i and once in A_j . Since there are $N - 1$ edges, this means that there are $2(N - 1)$ edge occurrences in the algorithm. But once a variable is calculated, it is not calculated again. Hence there are $N - 1$ calculations of variables with $N - 1$ bypasses, and the number of calculations is $O(N)$. The extension to the general case is obvious. It is also worth noting that for the case (4.14), the adjacency set is implicit in the data structure, as is of course the special cases of (4.1) and (4.3). The various aspects and nuances of the above are illustrated and clarified with the following example.

An Example

To clarify matters, consider a line $\ell \subset \mathbb{R}^9$ specified by two of its points whose coordinates appear in the first three columns of Table 4.1. As already discussed, a number of descriptions of the line in terms of linear equations involving pairs of variables is possible. The detailed construction of the chosen set of $\bar{\ell}_{i,j}$ is shown in Fig. 4.10. First the polygonal lines representing the two points P_1 and P_2 are drawn. Then the $\bar{\ell}_{i,j}$, corresponding to the choice of x_5 for *start variable*, are obtained by

Table 4.1. In the left three columns are the coordinates of two points specifying a line ℓ from which a set of eight $\bar{\ell}_{i,j}$ is constructed. The data structure used in the point construction algorithm, given in the next five columns, provides the $\bar{\ell}$ and their locations. Note the values $h_{i,j} = 1$ indicating finite and $h_{i,j} = 0$ ideal points. The points designating $x_k = \text{constant}$ (i.e., $k \in \tilde{S}$) have $i = x_{i,j} = 0$.

Coords	Point 1	Point 2	i	j	$x_{i,j}$	$y_{i,j}$	$h_{i,j}$
x_1	-4.0	2.0	1	5	2.4	-0.7	1
x_2	2.0	2.0	3	5	3.2	-0.3	1
x_3	-3.0	3.0	4	1	7.5	2.0	1
x_4	-1.6	2.0	8	5	5.5	0.25	1
x_5	1.5	-2.5	9	3	6.0	-1.0	0
x_6	-2.0	-2.0	0	2	1	2.0	1
x_7	-0.5	-0.5	0	6	5	-2.0	1
x_8	-1.0	3.0	0	7	6	-0.5	1
x_9	-4.0	2.0					

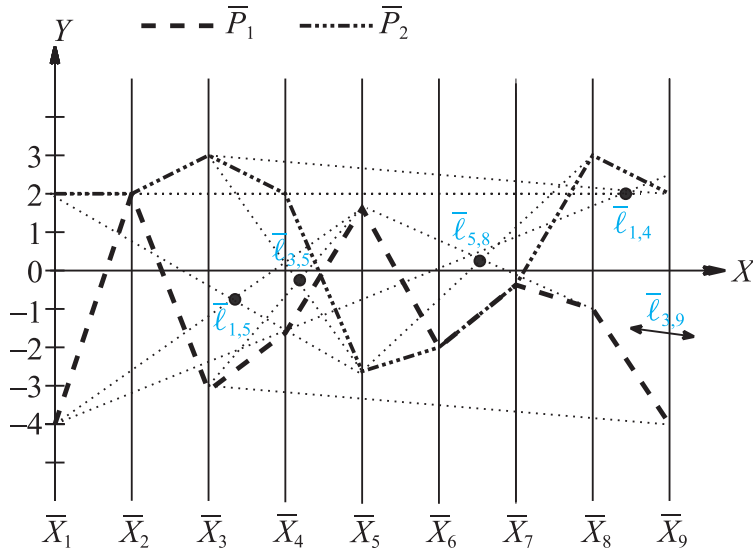


Figure 4.10. Construction of the $\bar{\ell}_{i,j}$ for the rooted tree in Fig. 4.9 and data in Table 4.1.

the intersection of pairs of thin dashed lines seen in the figure. For example, in order to obtain $\bar{\ell}_{1,5}$, the x_1 and x_5 coordinates of each point are joined. The intersection of

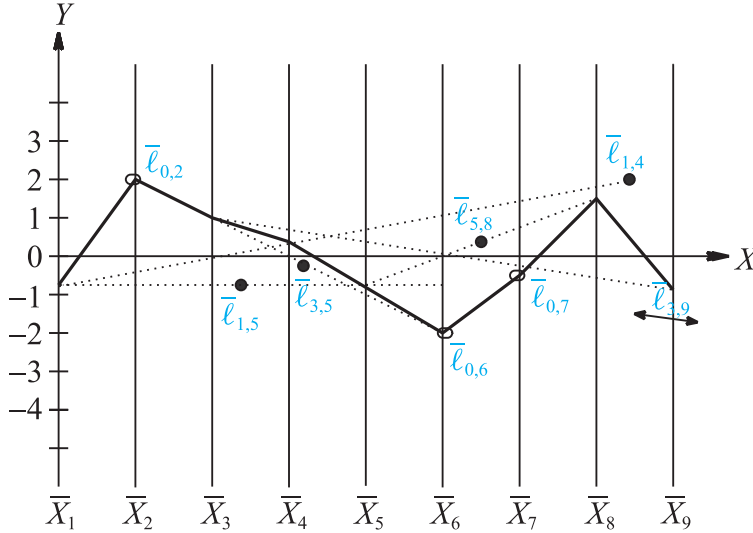


Figure 4.11. The $\bar{\ell}_{i,j}$ and the construction of a new point $P \in \ell$ for $x_5 = -0.75$.

these two lines is $\bar{\ell}_{1,5}$. The corresponding rooted tree for the *start* variable x_5 is the one already shown in Fig. 4.9. Altogether, for this tree the points required are $\bar{\ell}_{1,5}$, $\bar{\ell}_{3,5}$, $\bar{\ell}_{1,4}$, $\bar{\ell}_{5,8}$, and $\bar{\ell}_{3,9}$. The last is an ideal point, since $x_9 = x_3 = -1$; hence the slope is 1, and it is indicated in the figure as a direction. The corresponding data structure appears in the rightmost five columns of the table. Here the number of unconstrained variables is $M = 5$, and there are $N - M - 1 = 3$ constant variables.

The $\bar{\ell}_{i,j}$ together with the points $\bar{\ell}_{0,j}$, on the \bar{X}_2 -, \bar{X}_6 -, and \bar{X}_7 -axes, for the constant x_j are also shown in Fig. 4.11 together with the construction of a point $P \in \ell$ for a given value of x_5 . Recall that x_5 , the *start* variable, corresponds to the root of the tree. The construction can be easily modified for another choice of *start* variable, possibly $M = 5$.

Identifying the $\bar{\ell}_{i,j}$

There are certain practical problems that must be dealt with. One difficulty arises in labeling these indexed points. The following example illustrates the nature of the problem. Given the two points

$$P_1 = (a, a + b, \dots, a + (N - 1)b),$$

$$P_2 = (c, c + d, \dots, c + (N - 1)d),$$

where $b \neq d$ for any pair i, j , we have

$$\bar{\ell}_{i,j} = \left(\frac{a-c}{d-b}, \frac{ad-bc}{d-b} \right). \quad (4.24)$$

That is, *all* the $\bar{\ell}_{i,j}$ may be congruent (a situation that is revisited in Chapter 5). Even without such unusual contingencies, the labeling of the points may cause overcrowding in the figure. This suggests that a separate tabulation of the $\bar{\ell}_{i,j}$ may be preferable to attempting to label these points directly on the diagram.

Fun and Games with the ILM

set *Chapter = Multidimensional Lines, Subject = Multidimensional,*
Example = Indexed points.

The two points (polygonal lines) in 6-D specifying a line appear.

Make some of the coordinates of the two points equal.

Observe the resulting indexed points.

Again change the coordinates so that each polygonal line
 becomes a straight line.

Observe the coinciding index points.

4.1.6 Convenient Display of Multidimensional Lines

Lines with Positive Slopes

Often, the $\bar{\ell}$ are out of the scope of the picture. For such cases, two line segments (portions of the polygonal lines representing two points on the line) may be shown whose intersection (when extended) defines the $\bar{\ell}$. The question arises whether it is possible to recast the problem of line representation in such a way that “distant” points are eliminated. A scheme is proposed that not only eliminates ideal points from the representation but also has a number of other advantages.

Let us transform the variables by

$$x'_i = e_i x_i, \quad (4.25)$$

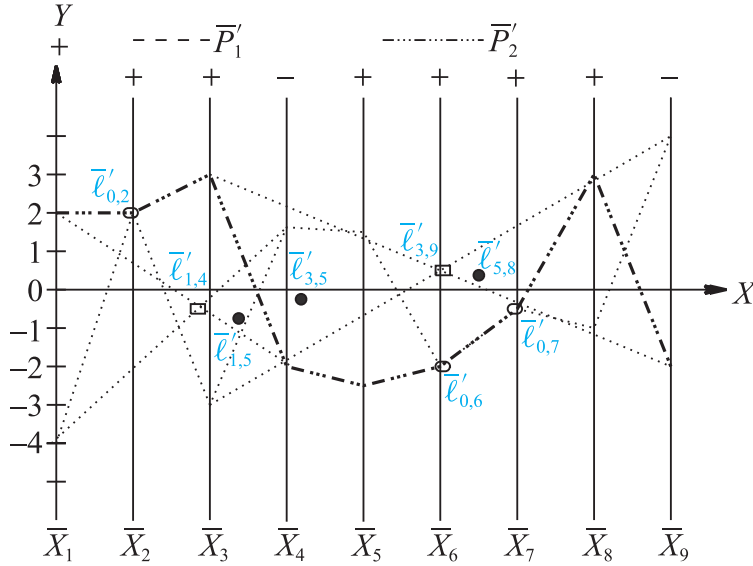


Figure 4.12. The $\bar{\ell}'_{i,j}$ for the example with the points \bar{P}'_1 and \bar{P}'_2 are drawn first.

Then the two points $\bar{\ell}'_{1,4}$, $\bar{\ell}'_{3,9}$, marked by rectangles, are constructed as indicated. All other points are unchanged, though a prime (') is added for notational consistency. Note the absence of ideal points.

where $e_i = +1$ for index $i \in \tilde{S}$. For the variables in S suppose that there is an edge on the tree joining the vertices x_i and x_j ; the vertex for x_j is either the root or lies between the vertex for x_i and the root. In this case, the linear equation is written as

$$x_i = m_i^j x_j + b_j.$$

The remaining e_i are developed recursively. Starting at the root index i , set $e_i = +1$ and visit the remaining nodes in order, setting $e_1 = +1$ or -1 to ensure that the slope m_i^j of the line

$$x_i = m_i^j x_j + b_j' \quad (4.26)$$

is negative, where $m_i^j = e_i e_j m_i^j$ and $b_i^j = e_i b_i^j$. Since the slope is negative, it follows immediately that the x -coordinate of $\bar{\ell}'_{i,j}$ is between $i - 1$ and $j - 1$, and that the y -coordinate is less in absolute value than the absolute value of b_i^j . With everything nicely bounded, all the $\bar{\ell}'_{i,j}$'s can be placed within the frame of a picture. Calculation of a point on the line proceeds as in the preceding algorithm for the

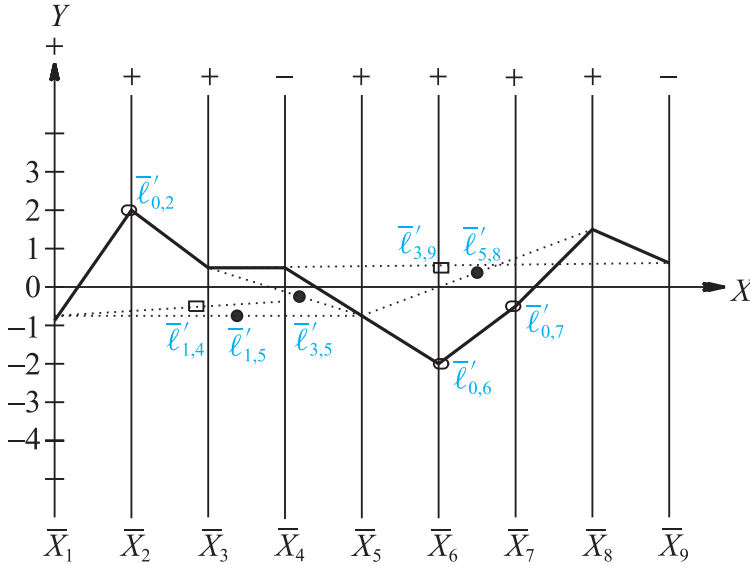


Figure 4.13. The polygonal line for the point P now denoted by P' in the new system \bar{X}'_i .

\bar{X} -system, followed by appropriate changes of sign and reflection in the \bar{X}_i -axis of those coordinates for which e_i is negative. Thus the algorithm is still of order $O(N)$. Likewise, construction proceeds with the \bar{X}' -system, followed by reflection in the \bar{X}_i -axis of those coordinates for which e_i is negative. However, while adjacency data depend only on the underlying graph structure, the numbers e_i are determined by the signs of the slopes, and hence must be recalculated for each case. Evidently, this does not alter the fact that the algorithm is of order $O(N)$. For the above example it turns out that the numbers e_i are successively 1, 1, 1, -1, 1, 1, 1, 1, -1. The data structure is given in Table 4.2, where now no fifth column is needed. It is really a simple matter to notice by inspecting Fig. 4.10 that only the points $\bar{\ell}_{1,4}$ and $\bar{\ell}_{3,9}$ do not lie between their corresponding two axes. Hence only for the indices 4 and 9 is $e_i = -1$, as indicated.

Whereas the transformation (4.25) $x_i \rightarrow x'_i$ works well for a single line, its application to several lines simultaneously can become problematic. Allowing different pairwise linear relations for each line greatly increases the prospect that their index-point representation can be brought within the framework of the picture (see exercises below). An elegant algorithm for the convenient display of *several* lines was obtained by S. Cohen-Ganor for his course project and is presented in the last chapter, entitled *Recent Results*.

Table 4.2. The $\bar{\ell}'_{i,j}$'s where the location of $x_{i,j}$ is now between the corresponding axes.

The $\bar{\ell}'$ are constructed from the transformed points \bar{P}'_i , $i = 1, 2$, and shown in Fig. 4.12. The previously constructed point P for $x_5 = 0.75$ is shown in the \bar{X}_j -coordinate system in Fig. 4.13.

i	j	$x_{i,j}$	$y_{i,j}$
1	5	2.4	-0.7
3	5	3.2	-0.3
4	1	1.875	-0.5
8	5	5.5	0.25
9	3	5.0	0.5
0	2	2.0	0
0	6	-2.0	0
0	7	-0.5	0

Exercises

- For the example of Section 4.1.5, choose a *start* variable other than x_5 and repeat the constructions. Namely,
 - exhibit the rooted tree,
 - construct and show the corresponding $\bar{\ell}$,
 - construct and show the point P' ,
 - construct and show the $\bar{\ell}'$,
 - construct and show the point P in the x'_i system.
- This is an example of the tree enumeration for $N = 4$. Show that,
 - only four distinct kinds of trees are possible, see Fig. 4.14, and
 - the numbers of distinct rooted trees (allowing permutation of vertex labels) are those shown below each tree.

How many distinct (not rooted) trees are there?

- The questions below pertain to the line representational problem of Section 4.1.6 and the transformation (4.25).
 - For $N = 3$ let ℓ and ℓ' be two lines such $m_1^2, m_2^3 < 0$ and $m_1'^2, m_2'^3 > 0$. Is it possible via the transformation of (4.25) to bring

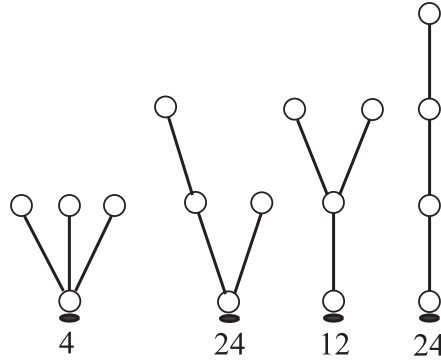


Figure 4.14. The four kinds of distinct rooted trees with labeled vertices (i.e., 1, 2, 3, 4). The number of trees allowing permutations of the vertex labels is listed below each tree type.

- all* the four $\bar{\ell}'$ between the corresponding axes? If not, give a counter-example.
- (b) For $N = 3$ and three lines ℓ, ℓ', ℓ'' state and prove necessary and sufficient conditions for the existence of a transformation $x_i \rightarrow x'_i$ such that the $\bar{\ell}_{i,j}, \bar{\ell}'_{r,s}, \bar{\ell}''_{u,v}$ are between the corresponding axes. Provide a specific example.
 - (c) Answer the previous question for specific numbers $N \geq 4$ and the number of lines $L \geq 4$.
 - (d) Generalize the result of the previous question for arbitrary N and L ; this is an open question, and likely to be *hard*.

4.1.7 ♣ FT-3 Rotations and Translations

The projective transformation dualities between parallel and Cartesian coordinates extend [23] to \mathbb{R}^N . Recall the duality between the rotation of a line ℓ about one of its points O corresponding to the translation of the point $\bar{\ell}$ on the line \bar{O} in \mathbb{R}^2 . Figure 4.15 shows a line $\ell \subset \mathbb{R}^5$ rotated about one of its points P to a new position ℓ' . This corresponds to the simultaneous linear translations of the $\bar{\ell}$ to new positions $\bar{\ell}'$ on \bar{P} . In fact, the pencil of *all* lines on P is represented by the collection of all quadruples $\bar{\ell}_{i,i+1}, i = 1, 2, 3, 4$, with $\bar{\ell}_{i,i+1} \in \bar{P}_{i,i}$, where $\bar{P}_{i,i+1}$ denotes the straight line of \bar{P} intersecting the \bar{X}_i and \bar{X}_{i+1} axes. Clearly both $\bar{\ell}_{i,i+1}$ and $\bar{\ell}'_{i,i+1}$ are on $\bar{P}_{i,i+1}$, and this leads nicely to the next topic.

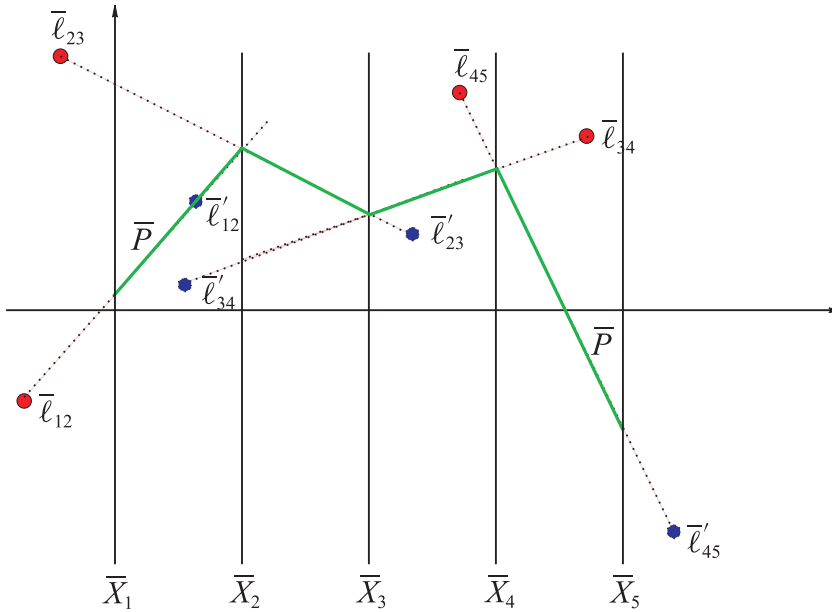


Figure 4.15. Two intersecting lines in 5-D.

Another interpretation of this picture is that the line ℓ rotates about one of its points shown here by the polygonal line \bar{P} . This corresponds to a linear translation of the $\bar{\ell}$ to new positions $\bar{\ell}'$ on \bar{P} .

4.2 Distance and Proximity Properties

4.2.1 Intersecting Lines

It is clear from Fig. 4.15 that $P = \ell \cap \ell'$, suggesting an easy intersection algorithm for the adjacent-variables description of lines. Given two lines ℓ, ℓ' we test for intersection by joining the $\bar{\ell}_{i,i+1}$ and $\bar{\ell}'_{i,i+1}$ with a line denoted by $\bar{P}_{i,i+1}$. Then ℓ and ℓ' intersect if and only if the line $\bar{P}_{i,i+1}$ intersect the x_{i+1} -axis at the same point as the line $\bar{P}_{i+1,i+2} \forall i$. The polygonal line so constructed represents the point where the two lines intersect. Formally, for two lines ℓ and ℓ' , described in terms of adjacent variables,

$$\exists P = \ell \cap \ell' \Leftrightarrow \bar{P}_{i,i+1} \cap \bar{P}_{i+1,i+2} = (i, p_{i+1}) \quad \forall i = 1, \dots, N-1, \quad (4.27)$$

where $x_{i+1}(P) = p_{i+1}$ the i th coordinate.

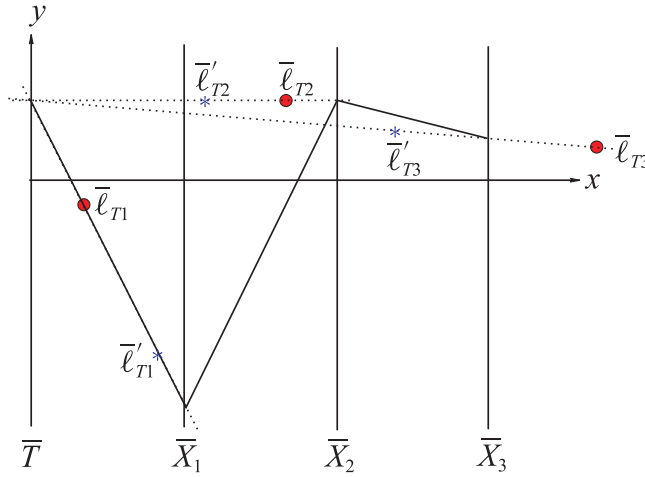


Figure 4.16. Intersection of two lines in 4-D for the base-variable line description. This shows the collision, space (x_1, x_2, x_3) , and time T coordinates of two particles moving with constant velocity whose trajectories ℓ, ℓ' are described by the $\bar{\ell}_{Ti}, \bar{\ell}'_{Ti}$.

Actually, $\bar{P}_{1,i} = (p_1, p_i)$ is the projection on the $x_1 x_i$ plane of the intersection point $P = \ell \cap \ell'$ for the base-variable description; Fig. 4.16. The intersection conditions for the base-variable description of the lines as given by

$$\left. \begin{aligned} \ell, \ell_{1i} : x_i &= m_i x_1 + b_i, \\ \ell', \ell'_{1i} : x'_i &= m'_i x_1 + b'_i, \end{aligned} \right\} \quad (4.28)$$

are

$$\exists P = \ell \cap \ell' \Leftrightarrow \alpha_i = -\frac{b'_i - b_i}{m'_i - m_i} = -\frac{\Delta b_i}{\Delta m_i} = p_1, \quad \forall i = 2, \dots, N, \quad (4.29)$$

where $x_1(P) = p_1$. Obtaining an intersection algorithm with $O(N)$ time complexity and generalizations to other line parametrizations are straightforward. There are special cases not covered by these conditions, whose treatment is simply not warranted at this stage (see [107] for a complete treatment and [97] for an earlier exposition).

4.2.2 Nonintersections

It is convenient to illustrate the situation in 4-D with the base-variable description of a line

$$x_i = v_i T + s_{o,i}, \quad i = 1, 2, 3. \quad (4.30)$$

For an interesting interpretation, let T denote elapsed time and (x_1, x_2, x_3) the space coordinates of a particle moving with constant velocity $\vec{V} = (v_1, v_2, v_3)$ and initial position $S_o = (s_{o,1}, s_{o,2}, s_{o,3})$. The complete trajectory information of the particle is provided by (4.30) and equivalently its three-point representation. The two sets of triple points $\bar{\ell}_{Ti}$ and $\bar{\ell}'_{Ti}$ describe the trajectories of two moving particles, and the construction in Fig. 4.16 shows an instance in which two particles collide; they go through the same point in space *at the same time*; a time-space intersection. Perhaps some of the power of the $\|$ -coords representation can be appreciated from this simple example.

Clearly, intersections in space (3-D and certainly higher) are very rare, so it behooves us to study nonintersections. The intersection algorithm is also useful to verify nonintersections as in illustrated in the sequence of Figures 4.17, 4.18, and 4.19 with the representing points of two lines in 4-D. Pairs of $\bar{\ell}_{Ti}$ and $\bar{\ell}'_{Ti}$ are connected with the three lines so formed not intersecting the T -axis at the same

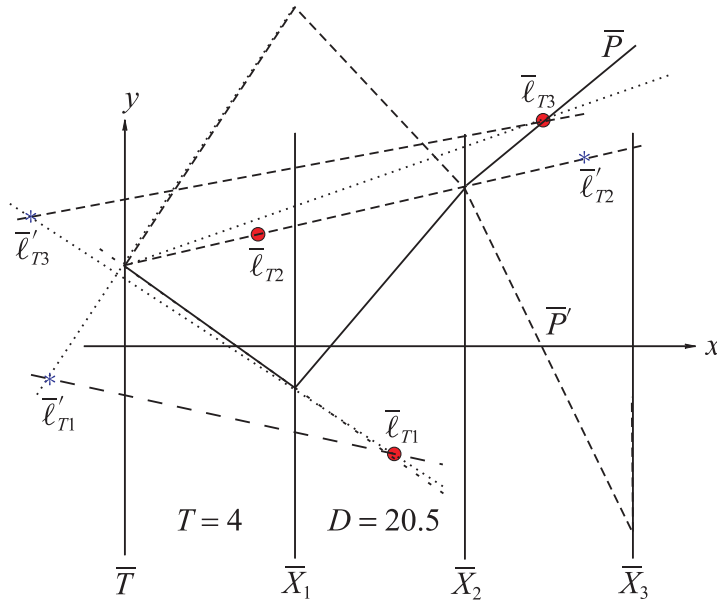


Figure 4.17. Nonintersection between two lines in 4-D.

Note the maximum gap on the \bar{T} -axis formed by the lines joining the $\bar{\ell}$'s with the same subscript. The polygonal lines represent the points P, P' , one on each trajectory, where the minimum distance of $D = 20.5$ occurs at time $T = 4$; Theorem 4.2.1.

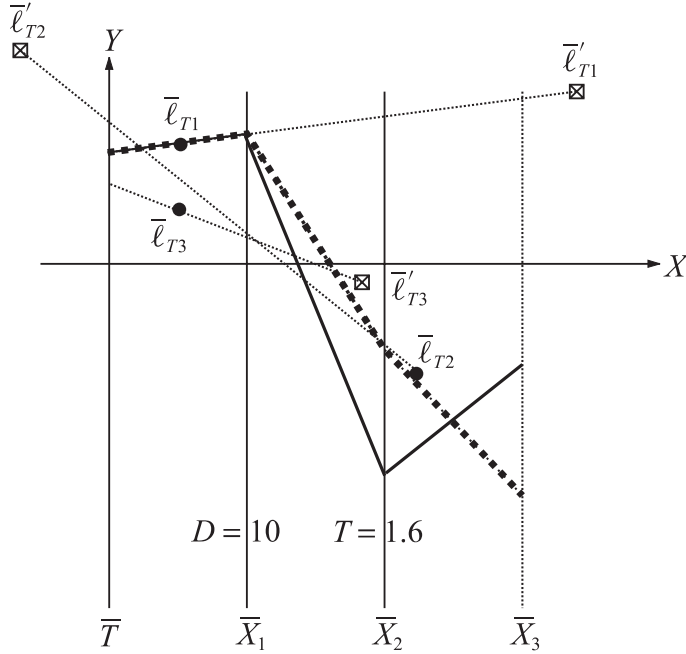


Figure 4.18. Another nonintersection between two lines in 4-D.

Here the minimum distance between the lines is 10, and this occurs at time = 1.6. Note the diminishing maximum gap on the \bar{T} -axis compared to that in Fig. 4.17. The polygonal lines represent the points where the minimum distance occurs.

value. The minimum distance \mathbf{D} between the lines is recorded. Denote the line on the points $\bar{\ell}_{T_i}, \bar{\ell}'_{T_i}$ by \bar{C}_i , its intersection with the \bar{T} -axis by

$$\alpha_i = \bar{T} \cap \bar{C}_i, \quad (4.31)$$

and the interval $I_{\bar{\ell}, \bar{\ell}'} = [\max \alpha_i, \min \alpha_i]$. It appears that $|I_{\bar{\ell}, \bar{\ell}}| \rightarrow 0$ as $\mathbf{D} \rightarrow 0$, which is reasonable, for we already know that $|I_{\bar{\ell}, \bar{\ell}'}| = 0$ when the two lines intersect ($\mathbf{D} = 0$). ♣ FT-3e

4.2.3 Minimum Distance between Two Lines in \mathbb{R}^N

Now we consider the more general problem of finding and, if possible, visualizing the minimum distance between two lines and the points on each line where the minimum occurs. In many problems what is required is to minimize the distance when one or more of the variables are *constrained* to the same value for both lines. An example is in air traffic control and motion planning in general, where of interest

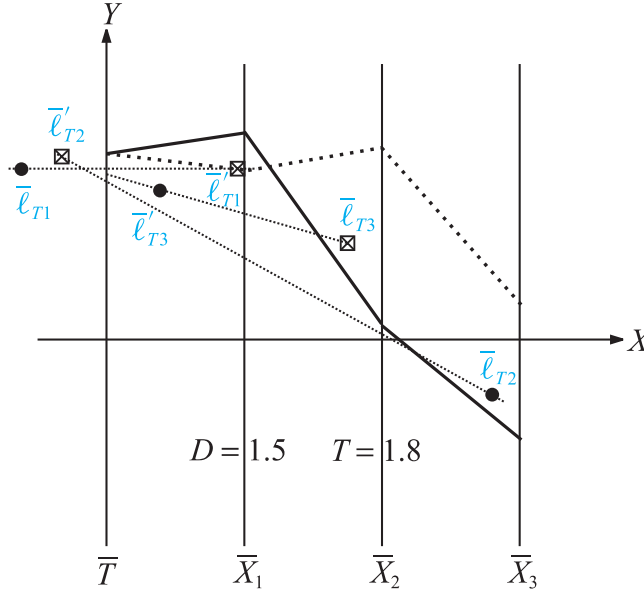


Figure 4.19. Near intersection between two lines in 4-D.

is knowing the *time*, and position, when two aircraft are the closest. Therefore, it is appropriate for such applications to constrain the *time* and find the minimum distance and position under that constraint. In that case, the minimum distance is zero when there is a time–space collision. For if time is not constrained, there will be two points $P_m = (t_m, x_1, \dots, x_{N-1}) \in \ell$, $P'_m = (t'_m, x'_1, \dots, x'_{N-1}) \in \ell'$ where the minimum distance occurs with $t_m \neq t'_m$ that would not be the closest distance between the two moving objects at the *same time*.

Constrained \mathcal{L}_1 Metric Minimum Distance

The following result suggests that in a way, the “natural” metric for \parallel -coords is the \mathcal{L}_1 (or “Manhattan”) metric, where the various components can actually be seen; see Fig. 4.20. The \mathcal{L}_1 distance between two points $P = (x_1, x_2, \dots, x_N)$ and $P' = (x'_1, x'_2, \dots, x'_N)$ is given by

$$\mathcal{L}_1(P, P') = \sum_{i=1}^N |x_i - x'_i|.$$

Now if $P \in \ell$ and $P' \in \ell'$ for the lines given by (4.28) and the base variable x_1 is constrained to be the same for both points, then the distance is given by

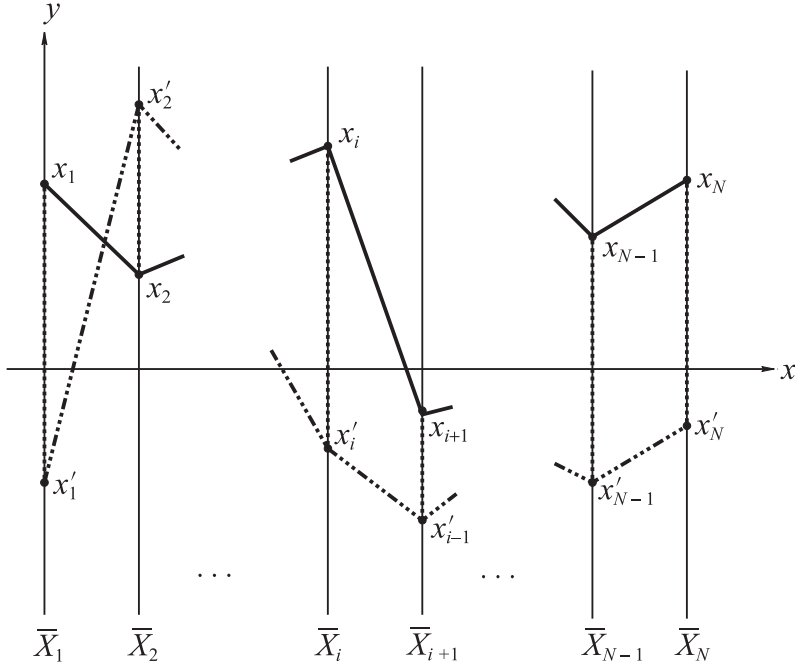


Figure 4.20. L_1 distance between the points $P = (x_1, \dots, x_i, \dots, x_N)$, $P' = (x'_1, \dots, x'_i, \dots, x'_N)$.

$$\mathcal{L}_1(x_1) = \sum_{i=2}^N |x'_i - x_i| = \sum_{i=2}^N |\Delta m_i| \left| x_1 + \frac{\Delta b_i}{\Delta m_i} \right| = \sum_{i=2}^N |\Delta m_i| |x_1 - \alpha_i|, \quad (4.32)$$

where the intercepts α_i are defined in (4.29).

Theorem 4.2.1. *The unique minimum value of $\mathcal{L}_1(x_1)$ is attained at $x_1 = \alpha_i$ for at least one $i = 2, \dots, N$.*

Proof. Step 1. On the interval $\alpha_k \leq x_1 \leq \alpha_{k+1}$,

$$\begin{aligned} \mathcal{L}_1(x_1) &= \sum_{i=2}^k |\Delta m_i| (x_1 - \alpha_i) + \sum_{k+1}^N |\Delta m_i| (\alpha_i - x_1) \\ &= \left(\sum_{i=2}^k |\Delta m_i| - \sum_{i=k+1}^N |\Delta m_i| \right) (x_1 - \alpha_i), \end{aligned}$$

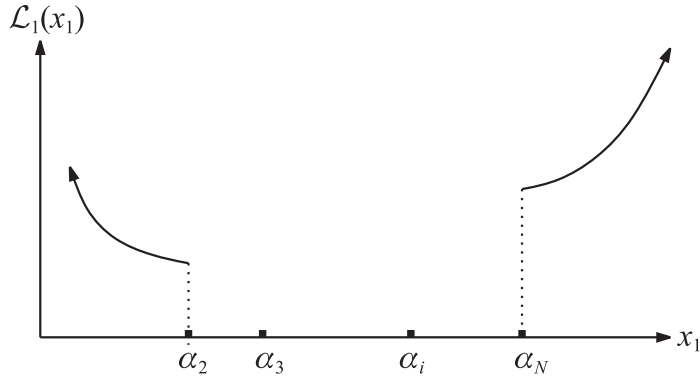


Figure 4.21. Monotone increasing portions of $\mathcal{L}_1(x_1)$.

is a linear function of x_1 with slope

$$T_k = \sum_{i=2}^k |\Delta m_i| - \sum_{i=k+1}^N |\Delta m_i|.$$

It attains its minimum at one of its endpoints unless $T_k = 0$, in which case $\mathcal{L}_1(x_1)$ is constant over the interval $[\alpha_k, \alpha_{k+1}]$.

Step 2. Reorder the index i if necessary so that $\alpha_i \leq \alpha_{i+1} \forall i$. First we consider the case $\alpha_i < \alpha_{i+1} \forall i$ and no $\Delta m_i = 0$. It is clear that for $x_1 \leq \alpha_1$, $\mathcal{L}_1(x_1)$ increases monotonically as x_1 decreases, since the $|x_1 - \alpha_i|$ increase $\forall i$. Similarly, for $\alpha_N \leq x_1$, again $\mathcal{L}_1(x_1)$ increases monotonically as x_1 increases, since the $|x_1 - \alpha_i|$ increase $\forall i$. So if $\mathcal{L}_1(x_1)$ has a minimum, it occurs at some $x_1 \in [\alpha_2, \alpha_N]$. However, this together with the conclusion of Step 1 implies that the minimum occurs at some $x_1 = \alpha_i$ for at least one value of i .

Step 3. Let $S_I = \mathcal{L}_1(\alpha_{I+1}) - \mathcal{L}_1(\alpha_I)$, and by the reordering in Step 2,

$$\begin{aligned} S_I &= \sum_{i=2}^N |\Delta m_i| |\alpha_{I+1} - \alpha_i| - \sum_{i=2}^N |\Delta m_i| |\alpha_I - \alpha_i| \\ &= \sum_{i=2}^I |\Delta m_i| (\alpha_{I+1} - \alpha_i) - \sum_{i=I+1}^N |\Delta m_i| (\alpha_{I+1} - \alpha_i) \\ &\quad - \sum_{i=2}^I |\Delta m_i| (\alpha_I - \alpha_i) + \sum_{i=I+1}^N |\Delta m_i| (\alpha_i - \alpha_I) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=2}^I |\Delta m_i|(\alpha_{I+1} - \alpha_I) - \sum_{i=I+1}^N |\Delta m_i|(\alpha_{I+1} - \alpha_I) \\
&= (\alpha_{I+1} - \alpha_I) \left(\sum_{i=2}^I |\Delta m_i| - \sum_{i=I+1}^N |\Delta m_i| \right).
\end{aligned}$$

Therefore,

$$S_I = (\alpha_{I+1} - \alpha_I) T_I. \quad (4.33)$$

Since $\alpha_{I+1} - \alpha_I > 0$, the sign of S_I is determined by the slope of the line segment over the interval (α_I, α_{I+1}) .

Step 4. By Step 1,

$$T_I = \sum_{i=2}^I |\Delta m_i| - \sum_{i=I+1}^N |\Delta m_i|, \quad T_{I+1} = \sum_{i=2}^{I+1} |\Delta m_i| - \sum_{i=I+2}^N |\Delta m_i|.$$

That is, T_{I+1} is found by moving the term $|\Delta m_{I+1}|$ from the negative to the positive sum. Therefore, T_I is monotone increasing with I . Hence from (4.33), if $T_I \neq 0$, the minimum of S_I is attained at a single point. Further, for $T_I = 0$, the minimum is attained over an entire interval (α_I, α_{I+1}) .

Step 5. We claim that the smallest value of I such that

$$\sum_{i=2}^I |\Delta m_i| \geq 0.5 \sum_{i=2}^N |\Delta m_i| \quad (4.34)$$

is the value of i for α_i where $\mathcal{L}_1(x_1)$ attains its minimum. For this is the first value of i for which the slope T_I can change sign. This observation provides an algorithm for finding the α_{*i} by construction. This provides a graphical construction for finding the minimum described in Fig. 4.22.

Step 6. When some of the α_i are equal, then by (4.33), there will be stationary values of \mathcal{L}_1 (i.e., regions where it is constant) and that may or may not be at the minimum; see Fig. 4.23.

Step 7. If there are $|\Delta m_i| = 0$ for i in a set E , then

$$\mathcal{L}_1(x_1) = \sum_E |\Delta b_i| + \sum_{\tilde{E}} |\Delta m_i| |x_1 - \alpha_i|, \quad (4.35)$$

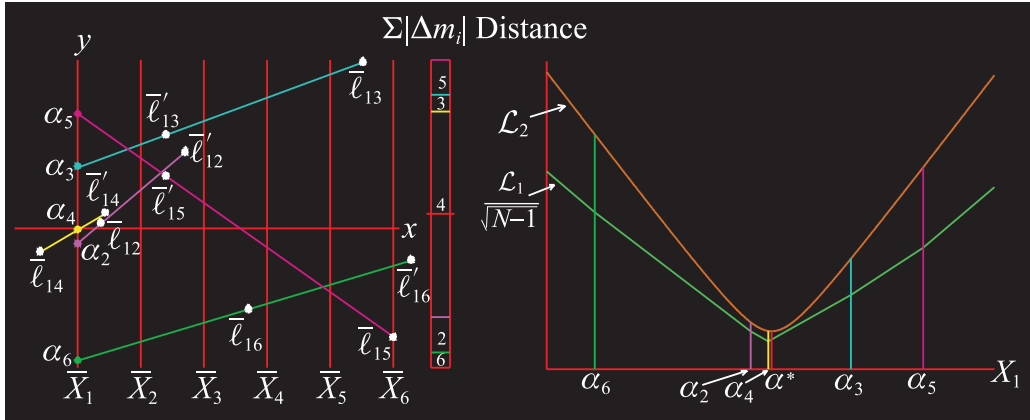


Figure 4.22. Constructing the $x_1 = \alpha_I$ minimizing the \mathcal{L}_1 distance between two lines.

For comparison, the minimum \mathcal{L}_2 distance occurs at $x_1 = \alpha^*$. The bar chart on the right of the \bar{X}_6 -axis shows the construction of the slopes T_I . There the $|\Delta m_i|$ are added in the order 6, 2, 4, 3, 5 obtained from the order of increasing α_i (as shown in the \bar{X}_1 -axis). The index I of the interval with the median value of the $\sum_{i=2}^N |\Delta m_i|$ provides the correct $x_1 = \alpha_I$. Here $|\Delta m_4|$ dominates the sum, yielding $I = 4$, so the minimum occurs at α_4 .

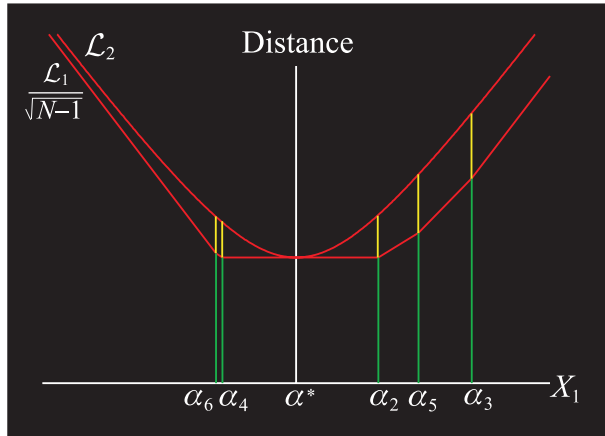


Figure 4.23. Here the \mathcal{L}_1 and \mathcal{L}_2 minimum distances are identical occurring at the same value of the constrained variable.

where E denotes the complement of E . This can be carried out by a modification of the above procedure. It turns out that for small $|\Delta m_i|$ and $|\Delta b_i|$ not small, the corresponding α_i are very large, and if not all the $|\Delta m_i|$'s are small, then the procedure simply ignores such values. ■

**** On the \mathcal{L}_2 Minimum Distance between Two Lines: Monotonicity with N**

Rather than the standard minimization approach, consider fitting a sphere with center at a point $P \in \ell$ and tangent to ℓ' at a point P' . The radius r of this sphere is the (minimum) distance between the two lines with P and P' the points where it occurs. To accomplish this fit, for any number α let $P \in \ell$, where $x_1(P) = \alpha$ and S is the sphere with radius r centered at P , i.e.,

$$S : \sum_{i=1}^N (x_i - m_i \alpha - b_i)^2 = r^2,$$

and for $P' \in \ell'$ with $x_1(P') = \alpha'$ and $x_i = x'_i$ for the x'_i of ℓ' (see 4.28), so we obtain

$$S \cap \ell' : \sum_{i=1}^N (m'_i \alpha' + b'_i - m_i \alpha - b_i)^2 = r^2.$$

We now “shrink” the sphere so that it is just tangent to ℓ' by finding the minimum r^2 . Expressing this in matrix form yields

$$r^2 = (\alpha \quad \alpha' \quad 1) \begin{pmatrix} M & -C & -D \\ -C & M' & D' \\ -D & D' & B \end{pmatrix} \begin{pmatrix} \alpha \\ \alpha' \\ 1 \end{pmatrix} \quad (4.36)$$

$$= M\alpha^2 + M'\alpha'^2 - 2C\alpha\alpha' + B + 2(D'\alpha' - D\alpha). \quad (4.37)$$

a quadratic relationship between the three unknowns α, α', r where

$$\begin{aligned} M' &= \sum_{i=1}^N m_i'^2, & M &= \sum_{i=1}^N m_i^2, & C &= \sum_{i=1}^N m_i m_i', \\ B &= \sum_{i=1}^N \Delta b_i^2, & \Delta b_i &= b'_i - b_i, \\ D &= \sum_{i=1}^N \Delta b_i m_i, & D' &= \sum_{i=1}^N \Delta b_i m_i'. \end{aligned} \quad (4.38)$$

The minimum r is found from

$$\frac{\partial r^2}{\partial \alpha} = 0 \Rightarrow M\alpha - C\alpha' = D, \quad \frac{\partial r^2}{\partial \alpha'} = 0 \Rightarrow C\alpha - M'\alpha' = D'. \quad (4.39)$$

The minimum is reached when

$$r^2 = B + D'\alpha' - D\alpha, \quad \alpha = \frac{M'D - CD'}{Q}, \quad \alpha' = \frac{CD - MD'}{Q}, \quad (4.40)$$

where $Q = MM' - C^2 \geq 0$ by the Cauchy–Schwarz inequality. When $Q = 0$ and $B \neq 0$, the lines are parallel and the constant square distance between them is $r^2 = B - D^2/M$. In \mathbb{R}^2 with $m_2 \neq m'_2$, the minimum is $r \equiv 0$, since the lines always intersect.

One Constrained Variable and Monotonicity with N

The minimization of the distance now follows for $x_1 = \alpha^*$ constrained to be the same for both ℓ, ℓ' . By a similar argument, s is the minimum distance in this case, and the α^* that minimizes it is found to be

$$\alpha^* = \frac{(D - D')}{V} = -\frac{W}{V} = -\frac{\sum_{i=2}^N \Delta b_i \Delta m_i}{\sum_{i=2}^N \Delta m_i^2}, \quad s^2 = B + (D - D')\alpha^*, \quad (4.41)$$

where $V = \sum_{i=2}^N (m'_i - m_i)^2$, $W = \sum_{i=2}^N \Delta b_i \Delta m_i$, or

$$s^2 = \sum_{i=2}^N \Delta b_i^2 - \frac{\left(\sum_{i=2}^N \Delta b_i \Delta m_i\right)^2}{\sum_{i=2}^N \Delta m_i^2}. \quad (4.42)$$

Applying Lagrange's identity yields the alternative form

$$s^2 \sum_{i=2}^N \Delta m_i^2 = \sum_{2 \leq k < j \leq N} (\Delta b_j \Delta m_k - \Delta b_k \Delta m_j)^2. \quad (4.43)$$

How does the minimum distance s vary as a function of the dimension N ? Let ℓ and ℓ' be two lines in \mathbb{R}^{N+1} with P and P' respectively their closest points and consider the projections ℓ_N, ℓ'_N on \mathbb{R}^N with closest points P_N, P'_N respectively. In (4.42), denote the distance s for \mathbb{R}^N by s_N and the corresponding α^* by α_N^* , $\Delta b_i = B_i$, $\Delta m_i = M_i$, and

$$D_N = \sum_{i=2}^N M_i^2, \quad C_N = \sum_{i=2}^N B_i M_i.$$

Subscript the analogous variables by $N + 1$ for \mathbb{R}^{N+1} . Then

$$s_N^2 = \sum_{i=2}^N B_i^2 - \frac{C_N^2}{D_N}, \quad \alpha_N^* = -\frac{C_N}{D_N}. \quad (4.44)$$

It is easily found that

$$s_{N+1}^2 - s_N^2 = \frac{(B_{N+1}D_N - C_N M_{N+1})^2}{D_N(D_N + M_{N+1}^2)} \geq 0,$$

and that $s_N \uparrow N$ (i.e., monotone increasing) with $s_N = s_{N+1} \iff$

$$\alpha_{N+1}^* = -\frac{\frac{C_N}{D_N} + B_{N+1}\frac{M_{N+1}}{D_N}}{1 + \frac{M_{N+1}^2}{D_N}} = -\frac{B_{N+1}}{M_{N+1}} = \alpha_N^*,$$

with $x_{N+1}(P) = x_{N+1}(P')$. All this shows that not only does s_N increase monotonically with N but also that in \mathbb{R}^N , the closest points between ℓ_N and ℓ'_N are *not* in general the projections of the closest points $P, P' \in \mathbb{R}^{N+1}$.

**** \mathcal{L}_1 versus \mathcal{L}_2 Constrained Minimum Distance**

For the constrained \mathcal{L}_2 distance on x_1 we consider the square

$$\mathcal{L}_2(x_1)^2 = \sum_{i=2}^N (\Delta m_i)^2 (x_1 - \alpha_i)^2.$$

It is left as an exercise to prove for the *unconstrained* case that in \mathbb{R}^N ,

$$\mathcal{L}_1(x_1) \geq \mathcal{L}_2(x_1) \geq \frac{\mathcal{L}_1(x_1)}{\sqrt{N}}. \quad (4.45)$$

Therefore, for the constrained case where the dimensionality is $N - 1$,

$$\mathcal{L}_1(x_1)^2 \leq (N - 1)\mathcal{L}_2(x_1)^2.$$

Hence if α minimizes \mathcal{L}_1 on a particular interval I , i.e.,

$$\mathcal{L}_1(\alpha) \leq \mathcal{L}_1(x_1) \quad \forall x_1 \in I,$$

and β minimizes \mathcal{L}_2^2 on I , i.e.,

$$\mathcal{L}_2(\beta)^2 \leq \mathcal{L}_2(x_1)^2 \quad \forall x_1 \in I,$$

then

$$\mathcal{L}_2(\alpha)^2 \geq \mathcal{L}_2(\beta)^2 \geq \frac{\mathcal{L}_1^2(\beta)}{(N-1)} \geq \frac{\mathcal{L}_1^2(\alpha)}{(N-1)}. \quad (4.46)$$

From (4.45) and (4.46) we obtain upper and lower bounds for the minimum $\mathcal{L}_2(x_1)$ in terms of the minimum of $\mathcal{L}_1(x_1)$. It turns out that these bounds are tight, since there exist cases in which these minima are equal for the same x_1 . Still, it seems that the \mathcal{L}_2 minimum distance does not lend itself easily to visualization in $\|\cdot\|$ -coords as the \mathcal{L}_1 distance does. Consider this a challenge!

Finding the minimum distance between lines raises the question of how to measure the “closeness” — according to some criterion — between lines. This is a matter of considerable interest in various applications such as computer vision and geometric modeling. Gauging the “closeness” between things of the same kind is really the subject matter of *topology*, whose original name was *analysis situs*. Later, in Chapter 8, various topologies for lines, p -flats, and hyperplanes are studied in depth.

Exercises

1. Write an algorithm for constructing a point on a line given the $\bar{\ell}$ and state its complexity.
2. Write an algorithm for constructing the representation of a line given two points P_1, P_2 on the line. Be careful and delimit the special cases.
3. Write an algorithm for constructing the intersection of two lines or verifying nonintersection and state its complexity. Clarify the special cases.
4. Write an algorithm for constructing the constrained \mathcal{L}_1 minimum distance and state its complexity. List all special cases in which the algorithm fails and why.
5. Prove (4.45). Hint: use induction on the dimension N .
6. Construct a case in which $\mathcal{L}_1(x_1)$ has stationary portions. State conditions for this to occur in general.
7. Construct a case in which the $\mathcal{L}_2(x_1)$ and $\mathcal{L}_1(x_1)$ minima are equal for the same x_1 .
8. Write an algorithm for minimizing (4.35) and state its complexity.

4.2.4 ** Air Traffic Control

Around 1985, the interest generated by the impending design and construction of a new air traffic control (ATC) system in the United States led, among other things, to exploration for new information displays. One of them was based on the $\|\cdot\|$ -coords methodology (US patent # 4,823,272), and it is briefly illustrated here in this,

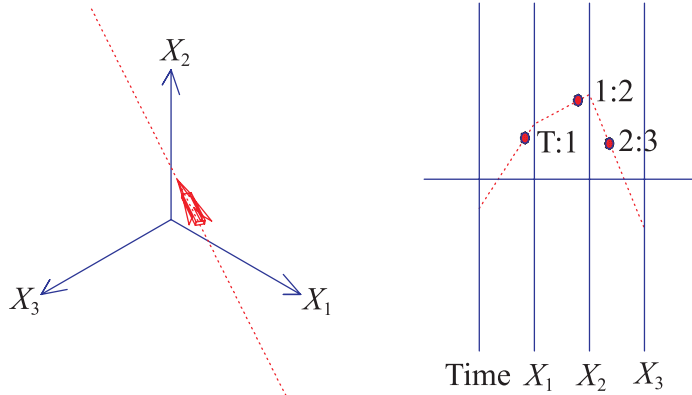


Figure 4.24. Path (left) and the full trajectory information (right) of an aircraft. In \parallel -coords, the aircraft's position can be found and displayed at any *time*.

the concluding section of the chapter. The reader interested on some nontechnical background on ATC is referred to [76] and for an early description of the then projected new system AAAS to [96].

Displaying Trajectory Information

The trajectory of an aircraft (considered as a point) with constant velocity, as occurs for much of the time in civil aviation, is a line segment in 4-D, which may be described by

$$\begin{cases} x_1 = v_1 T + s_1, \\ x_{i+1} = v_2 x_i + s_2, \\ x_{i+1} = v_3 x_i + s_3, \end{cases} \quad (4.47)$$

where the x_i are the space coordinates, T is time, $\mathbf{V} = (v_1, v_2, v_3)$ is the velocity vector, and $S = (s_1, s_2, s_3)$ is the initial position at $T = 0$. We already know that in parallel coordinates the complete trajectory information is given by three *stationary* points, which may be labeled²¹ **T:1**, **1:2**, and **2:3**. Using the \bar{T} (time)-axis as a clock, the present position of the aircraft as well as where it was and where it will be at any given time can be found and displayed as in Fig. 4.24. Recall that the horizontal coordinate of the points representing the trajectory given by (4.47) is determined only from the quantities m_i , with m_1 the velocity component in the x_1 direction.

²¹For brevity we omit the $\bar{\ell}$ part from the notation whenever it is clear from the context.

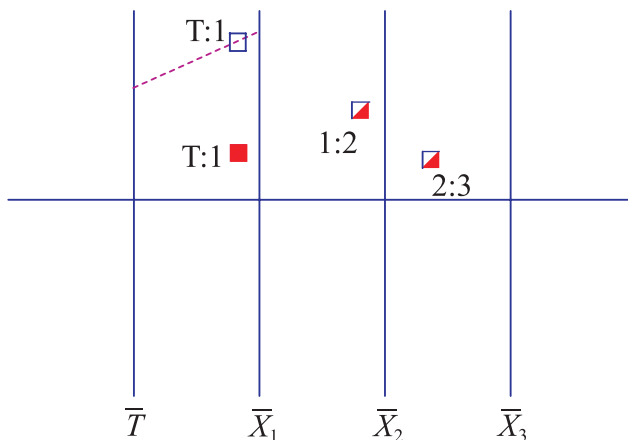


Figure 4.25. Two aircraft flying on the same path, since their **1:2** and **2:3** points coincide. They have a constant separation, with their velocities the same, since **T:1** points have the same x -coordinate.

With x_3 the altitude, m_2 is the tangent of what in aviation parlance is called the *heading*, and the horizontal position of **T:3** provides the vertical velocity.

Certain properties of the trajectories are immediately evident from the display. Consider the situation in which aircraft are flying on the same *path*, as is often the case in civil aviation. An example is shown in Fig. 4.25, where the points **1:2** and **2:3** are shared, since these are the points describing the path in \parallel -coords. Further, it is evident that they have the same velocity, since the points **T:1** have the same horizontal coordinate. Otherwise, if one point is between the \bar{T} , \bar{X}_1 and the other outside axes, their velocities must have opposite signs, and since they have the same path, they are flying in opposite directions. If they are both between the axes or both outside the axes, then the leftmost point corresponds to the greater velocity. Other properties are also “visible.”

The time and positions at which the distance between two aircraft is minimum is displayed in Fig. 4.26. This critical situation is accurately portrayed in \parallel -coords, while in 3-D displays, not to speak of 2-D projections, it can appear misleading. The possibility of transforming angular to lateral deviations is illustrated in Fig. 4.27. This is important since our eyes can distinguish lateral deviations better rather than angular deviations, such as those that may occur in the assigned path. This is a “human factors” advantage which may be partially offset by the nonlinear scale, a result of the denominator $(1 - m_i)$ of the points that represent the path.

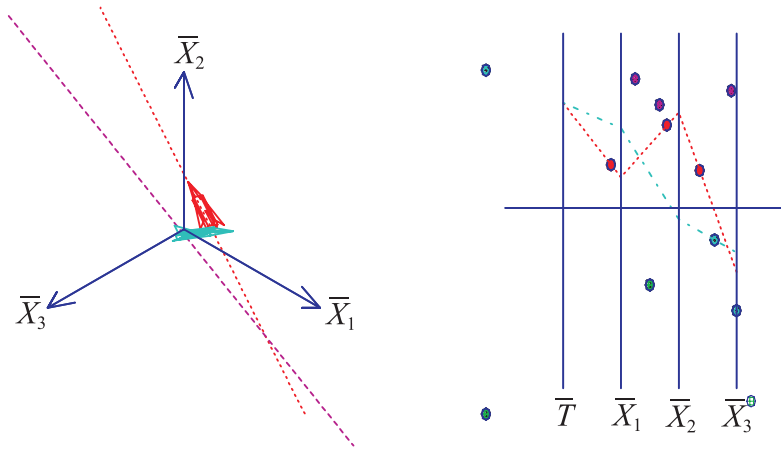


Figure 4.26. Closest approach of two aircraft.

The time and closest positions are clearly seen in \parallel -coords. Appearances can be misleading in a 3-D (near perspective) display, where the aircraft appear to be nearly colliding. It is even more uninformative in 2-D, where only a projection is displayed.

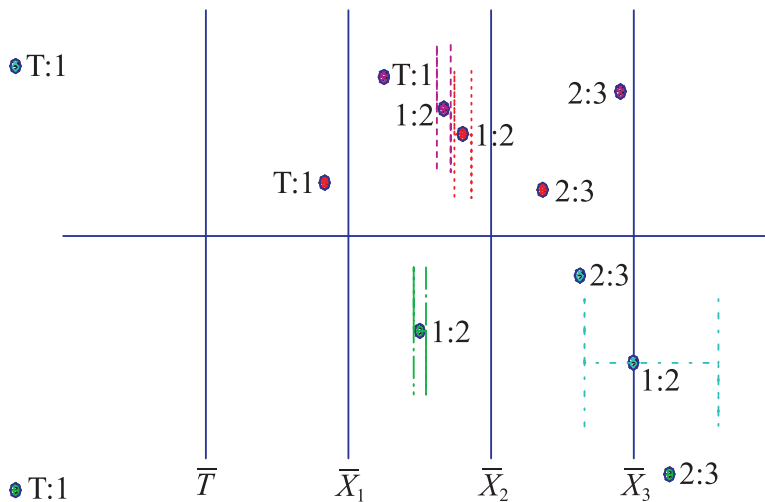


Figure 4.27. Transforming deviations in heading (angle) to lateral deviations.

Particle Fields, Scrapes, and Conflic Intervals

The goal of ATC is to direct aircraft safely to their destination while minimally interfering with their planned trajectories. This involves maintaining a minimum

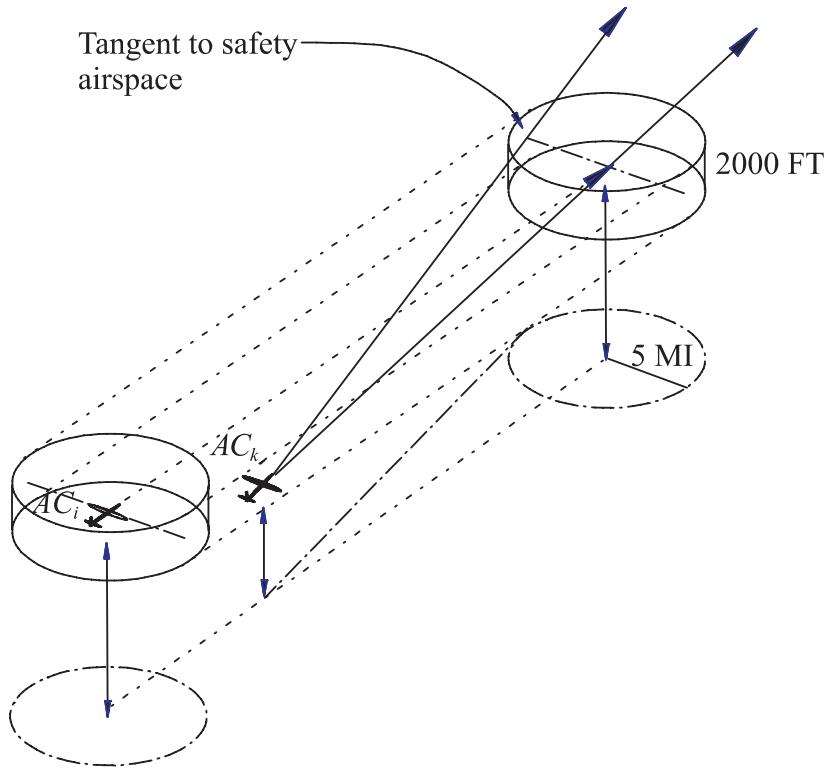


Figure 4.28. Protected Space in 3-D is a thin cylinder.

separation between the aircraft; detecting impending *conflict*, which are violations of the minimum separation; and *resolving* the conflicts by changing the trajectories subject to certain constraints and priorities. The typical safety requirements in 3-D (sometimes referred to as “separation standard”) involve horizontal and vertical separations, as shown in Fig. 4.28. The protected airspace is a thin cylinder usually 2,000 feet high and diameter of 5 nautical miles. This shape is sometimes called a “puck,” due to its similarity with the hockey puck. Conflict resolution is an instance of the theoretical *asteroid avoidance problem* (AAP), which turns out in general to be NP-hard [21]. In ATC, for a number of good reasons it is preferred to resolve conflicts if possible without altitude changes. An algorithm for planar (i.e., no altitude change) conflict detection and resolution in ATC (USA patent # 5,058,024) was proposed in conjunction with the design competition for the new ATC system (AAAS). It was tested on some complex groupings of flying aircraft (“complex scenarios” for short) that were provided by the USA Federal Aviation

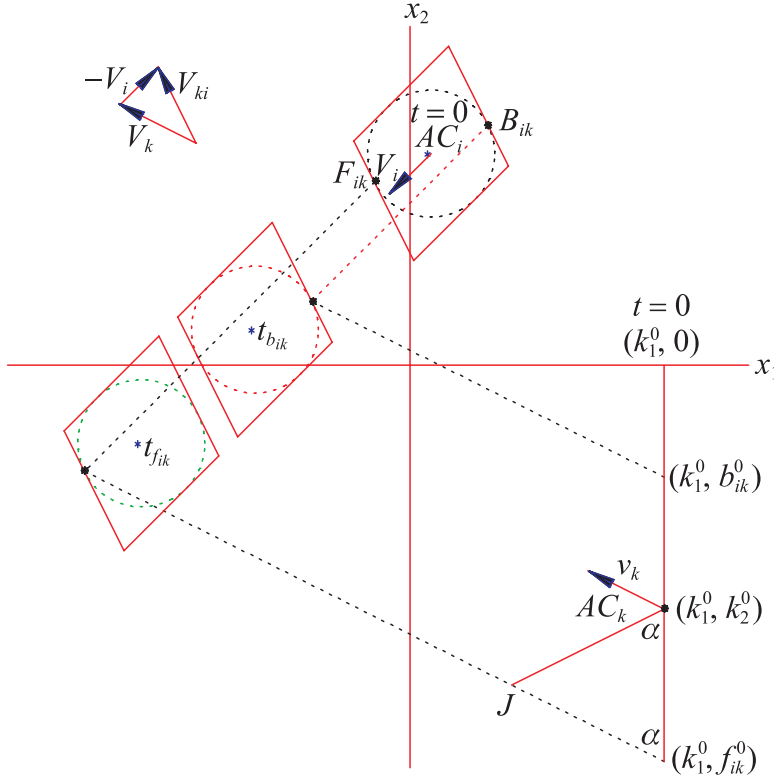


Figure 4.29. Relation between maneuver speed and turn angle.

An equal-speed maneuver results with turn angle α , whereas turn angles greater than or less than α require speeds slower or faster than $|\vec{V}_k|$ respectively.

Administration [127]. We shall give the rudiments of the algorithm and then apply it to one of the complex scenarios.

Given a set of aircraft considered as points moving with constant velocity, if the minimum separation to be maintained is denoted by d , then the protected airspaces, for the planar case, are circles with radius $d/2$ centered at each aircraft and moving with the same velocity. The ensuing conflict and resolution analysis is done first for a pair of circles and then for an arbitrary number of circles. There is no loss of generality in reducing this to the point-circle configuration above (by shrinking one circle to) a point (the aircraft itself) AC_k with velocity \vec{V}_k , while the circle C_i for the aircraft AC_i has double the original radius and velocity \vec{V}_i , as shown in Fig. 4.29. Here the velocities are assumed to be constant, as when the aircraft are flying after having reached their cruising altitude. A line ℓ , taken here for convenience to be vertical, going through AC_k is shown at the initial time $t = 0$. On ℓ a vector f_{iel} is

defined with the points on the line now being considered as “*particles*” endowed with the same velocity \vec{V}_k as AC_k for $t \geq 0$. It is clear that there exists exactly one such particle, initially in position (k_1^0, b_{ik}^0) , that just “scrapes” (i.e., is tangent to) the circle C_i at time $t_{b_{ik}}$. Similarly, there exists a unique particle initially in position (k_1^0, f_{ik}^0) that exactly scrapes C_i from the front. The particles starting above b_{ik}^0 pass the circle safely from the back, particles below f_{ik}^0 pass in the front, while those in the interval $I_{ki} = [f_{ik}^0, b_{ik}^0]$ eventually intersect the circle. This provides the conflict *detection*, since for any aircraft on ℓ , only those on I_{ki} will be in conflict with AC_i . For lack of a better name, I_{ki} is called the *conflic interval of i with respect to k* . There is more information in Fig. 4.29 worth pointing out. The parallelogram enclosing C_i has two sides parallel to \vec{V}_i and two parallel to \vec{V}_{ki} , the velocity of k relative to i . That is, an observer sitting on C_i will see the particles passing by with velocity \vec{V}_{ki} . Only the particles entering this parallelogram will intersect C_i , and these are the particles on I_{ki} . So the particle scraping the circle from the back will be seen by such an observer traveling on the top (back) tangent to C_i and touching it at the point B_{jk} . That, in fact, is the point where the back scrape will occur. The intersection of this side with ℓ occurs precisely at the point (k_1^0, b_{ik}^0) (why?), providing the particle responsible for the back scrape. Since this particle has velocity \vec{V}_k , the intersection of a line in this direction with the path of B_{jk} (which has velocity \vec{V}_i) must be the position where the back scrape occurs, from which, in turn, the time $t_{b_{jk}}$ can be found. All this is easily seen by noticing that the triangle so formed is similar to the triangle showing vector subtraction in the upper left. The situation is exactly the same vis-a-vis the front scrape. In this way, running through all the indices $i = 1, \dots, N, i \neq k$, the corresponding conflict intervals can be found.

In \parallel -coords the paths of the particles on I_{ki} are transformed into points. These paths all being parallel, they are represented by points and will, in turn, correspond to an interval, say \tilde{I}_{ki} on a vertical line at $x = \frac{1}{1-m}$ with m the slope of \vec{V}_k . The situation becomes more interesting when this is considered $\forall i = 1, \dots, N, i \neq k$. Since m is still the same, all the corresponding \tilde{I}_{ki} fall on the *same* vertical line. Let us see how this works on one of the complex scenarios (# 8) from [127]; where 10 aircraft are involved. What makes it “interesting” is that 6 of these, shown in Fig. 4.30, are flying at the same altitude and somewhat menacingly toward each other; several impending conflicts are seen in Fig. 4.31. The corresponding conflict intervals are shown in Fig. 4.32, showing the paths for the **Back** scrapes and **Front** scrapes of the aircraft # 2, \dots , 6 and the times when they occur. The times are computed from (4.47) or directly from its point representation in \parallel -coords.

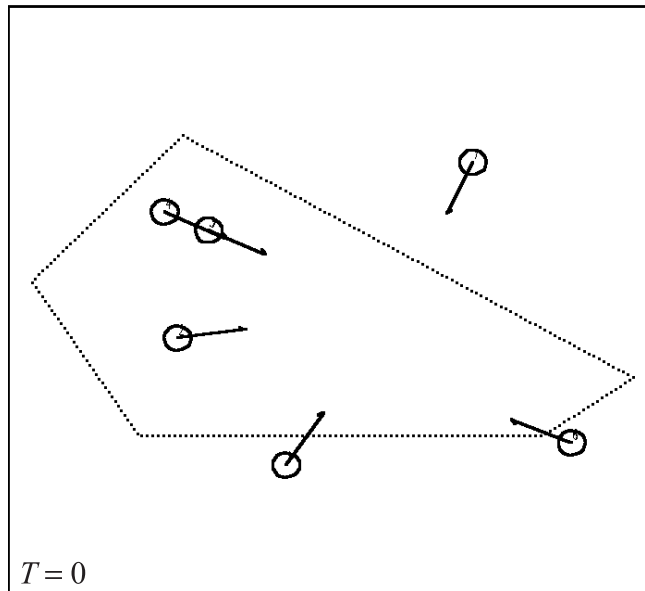


Figure 4.30. Six aircraft flying at the same altitude.

These positions are at a certain time (taken as 0 seconds and shown on the left-hand-corner). Circles centered at each aircraft are the protected airspaces, with the diameter the minimum allowable separation. The arrows represent the velocities.

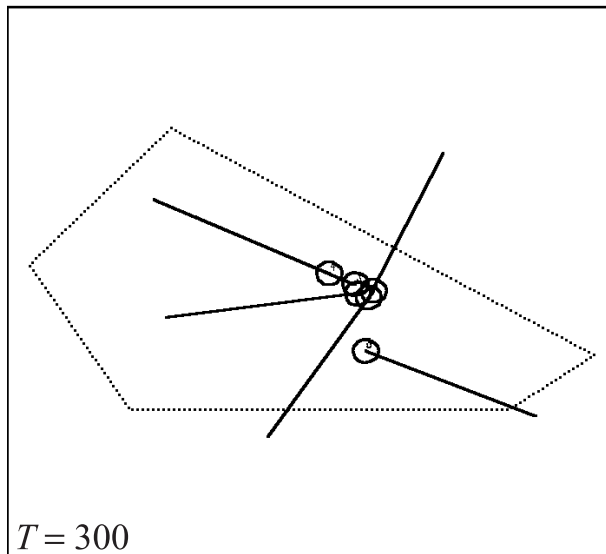
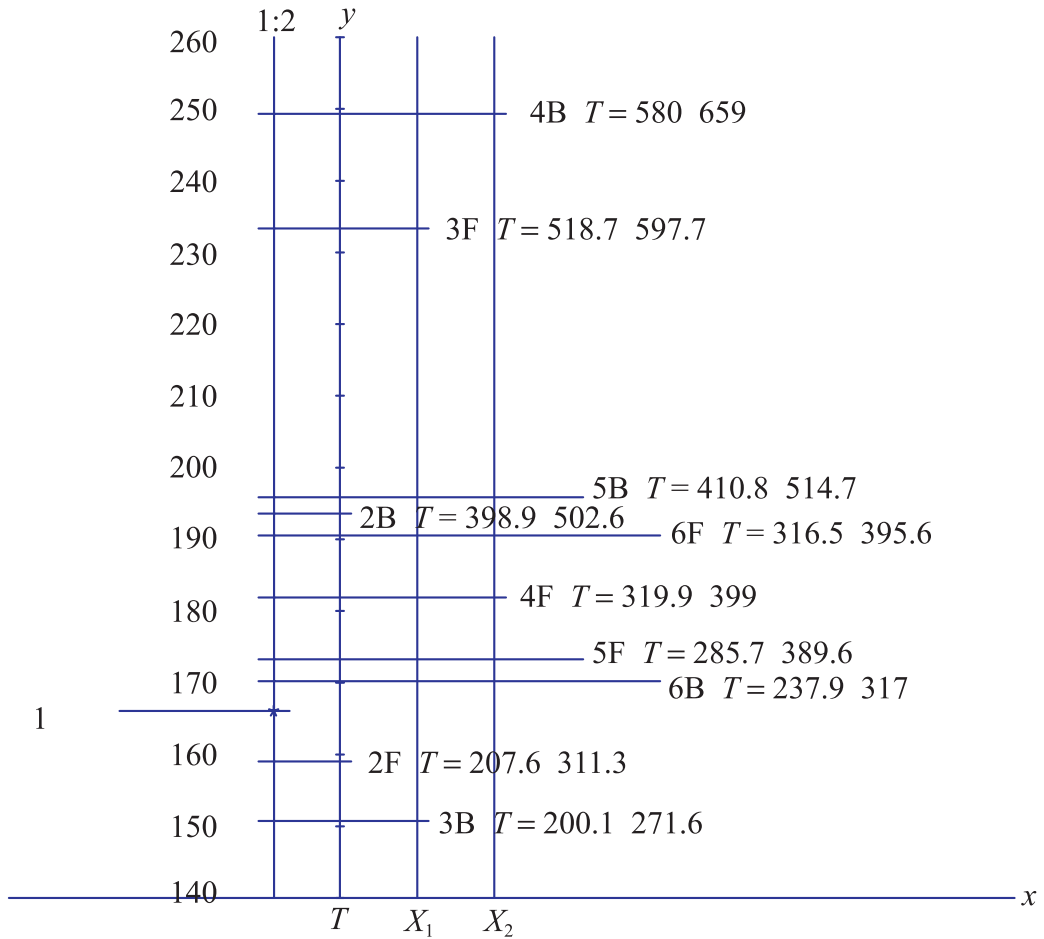


Figure 4.31. Conflicts, indicated by overlapping circles, within the next five minutes.



Aircraft 1 versus the remainder

Figure 4.32. Conflict Intervals I_{1k} , $k = 2, 3, 4, 5, 6$.

These are with respect to aircraft # 1 for the complex scenario shown in Fig. 4.30 and the conflicts shown in Fig. 4.31. The path of aircraft # 1 is shown as a point on the vertical line. Lowest point shown is **3B** and represents the path of **Back** scrapes # 3, and next to it is the *time* when this occurs. In general, **B** or **F** denotes back or front scrapes for the indicated aircraft number and the time at which the scrape occurs.

Conflic Resolution

Now the fun begins! Let us denote by \bar{k} the point representing the path of AC_k . Clearly, AC_k is conflict with aircraft # $i \Leftrightarrow \bar{k} \in I_{ki}$. For example, from Fig. 4.32 it is easily seen that aircraft # 1 is in conflict with # 2 and # 3 and no others. The

challenge is to resolve the conflicts with the allowable maneuvers, which in our case are *parallel offsets*, which are commonly used en route. Such a maneuver consists of:

1. a turn (left or right) that is bounded above, say, by a θ_{MAX} , that is, a turn that is relatively “small,”
2. followed by a relatively short straight path; this is the “offset,”
3. then a second turn returning the aircraft to a path parallel to the original one (i.e., the same heading). The new path is not allowed to be “too far” from the original, i.e., the distance between the path is bounded above, say, by s .
4. At all times, the aircraft’s speed, i.e., $|\vec{V}|$, is constant, that is, no slowdowns or speedups are allowed even during the maneuvers.

The turn here is idealized in the sense that it is considered “instantaneous” in time and occurs at a point in space. All this can be adjusted to more realistic models based on the capabilities of various airplanes, policies for various airlines, and air sectors, as well as many other constraints and priorities such as time-to-conflict (“validation time”), all of which are beyond our scope here. The strategy for conflict resolution then is to use parallel offset maneuvers subject to the given constraints in order to redirect the aircraft to conflict-free paths. The conflict intervals are well suited for this task. Since eventually the new path is parallel to the original path, the idea, when # j is in conflict with some aircraft, is to find the nearest conflict-free path, say \bar{j}' , and check to see whether under the constraints, the maneuvers allow # j to reach this path *prior to the scrape*. That is, we want # j to behave like the particle originally with path j' in a neighborhood of the scrape. Specifically, for # 1 we see in Fig. 4.32 that a good candidate is the path of the particle causing the back scrape **3B**. This is found by taking the union of all conflict intervals containing $\bar{1}$, which is also an interval (from a basic theorem in topology) and examining the closest endpoint. In general, the new interval’s endpoint may lie on conflict intervals not included in the union, such as **3F** $\in I_{14}$, in which case these intervals are added to the union until a conflict-free path is found. This process terminates, since there is a finite number of intervals. Then the closest endpoint is checked to see whether it is *feasible*, that is, whether the aircraft in conflict, such as # 1 here, can be redirected to the new path subject to the constraints. It is not difficult to apply the maximum-offset constraint.

To see how equal-speed maneuvers can be achieved, refer again to Fig. 4.32, where the turn-angle-to-maneuver-speed relation is illustrated. When AC_k makes a turn with angle α , an isosceles triangle is formed such that after the turn, the aircraft traveling with speed $|\vec{V}_k|$ arrives at J at the *same time* as the particle originally at f_{ik}^0 , since they both travel with the same speed over an *equal distance*. Hence after

the point J , AC_k , now in its new path, mimics the behavior of the particle, namely the scrape, in the neighborhood of some circle. From this picture we can also see that turns of more or less than α will reach the path of f_{ik}^0 sooner or later than the particle, necessitating a slowdown or speedup by AC_k to accomplish the same scrape as the particle. Of course, a choice of lines other than the vertical provides more and different options, and this is exploited in the implementation.

The general AAP is NP-hard, so even its specializations may have very high complexity. Since conflict resolution requires a real-time solution, one way to handle the need for fast response is by *cascading* the algorithm in various levels, which we briefly outline below, of increasing complexity and power. Starting with the first level (having the lowest complexity), the set of aircraft is processed by successive levels until either a resolution is found or no resolution is found by the highest level, though, in our experience, satisfactory resolutions were found at the lowest level. It is, of course, even better to run the algorithms for the various levels in parallel. The algorithm is *greedy*, and in testing, we found that different resolutions are found by the various levels. Different resolutions (or no resolutions) may also be found by reordering aircraft on input. Listing the aircraft in order of decreasing number of conflicts they are involved in turns out to be a “good” ordering. Other ordering heuristics, which also consider the time-to-conflict, have been tried with varying success. When several resolutions for a conflict scenario are available, rating them as to their “desirability” has turned out to be a very elusive and very interesting problem.

In the simplest case (“simple rules”), an aircraft in conflict, say AC_j , is placed on the path found, as indicated above, satisfying all the constraints. This is repeated with the *updated* (due to the repositioning of AC_j) conflict intervals. If this works out for all the aircraft in the input list, a resolution is found; otherwise, the next level is invoked. The computational worst-case complexity for this level is $O(N^2 \log N)$, where N is the number of aircraft.

When the simple rules level finds no resolution for a particular aircraft AC_k , a sweep of the conflict intervals determines which aircraft prevented it (i.e., blocked it) from being resolved, and there may be more than one. A criterion is applied in order to choose the “most troublesome” aircraft blocking the resolution, temporarily removing it and resolving the conflicts, if possible, using the simple rules. Such potential conflict resolution is provisionally taken, and a resolution for the remaining aircraft, including the one that was removed, is attempted. If successful, the resolution spawns additional maneuvers for the blocking aircraft one at a time. This backtracking can be nicely implemented with recursion.

The worst-case complexity of the recursion is exponential. In order to terminate, the maximum number of times R that the removal of a blocking aircraft is

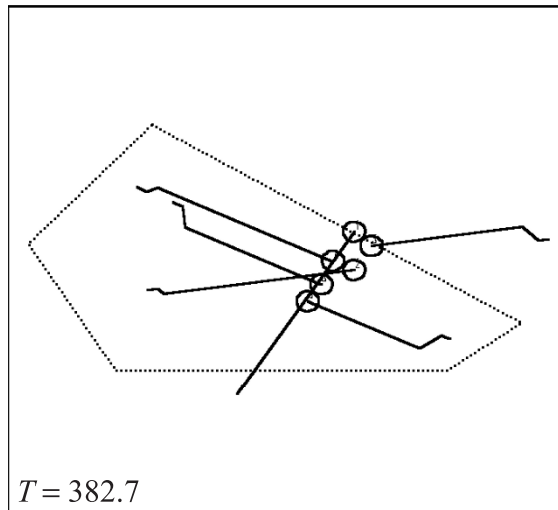


Figure 4.33. Conflict resolution with parallel-offset maneuvers. Three pairs of tangent circles.

allowed to take place is fixed a priori, yielding that the maximum complexity is $O(N^{R+1} \log N)$. It was also found profitable to experiment with different reordering heuristics rather than invoke backtracking. Though frequently successful, no clear scenario-independent criteria for the reordering were found. However, using several particle lines rather than one led to resolutions for the vast majority of scenarios tried using only the simple rules. For equal-speed maneuvers, every particle line corresponds to a specific maneuver turn. In addition to the fact that “idealized” turns are used here, we must also consider that a pilot cannot execute a turn with absolute precision. So a certain tolerance, say θ , for the error is involved. So placing the particle lines in multiples of θ and covering in this way the allowable turn range (i.e. $\pm\theta_{\max}$) is a way to check for all *implementable* resolutions without an increase in the time complexity.

Let us illustrate all this with the chosen scenario shown in Fig. 4.30. To estimate its difficulty in a practical way, an expert air traffic controller was asked to resolve it. He could resolve only four of the aircraft without altitude change. Even that required a great deal of time, and the solution was certainly not “real-time.” Our algorithm was able to resolve all conflicts without altitude changes using only two particle lines, the resolution being shown at two time frames in Figs. 4.33 and 4.34. There are several instances of tangent circles. This is because the algorithm is built to utilize the scrapes, which, in turn, result in minimally disturbing the aircraft from their original paths. The figures should illustrate the difficulty in finding constrained

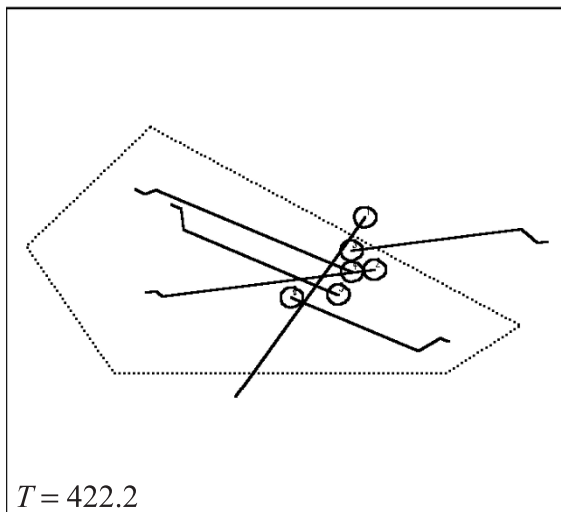


Figure 4.34. A triple tangency.

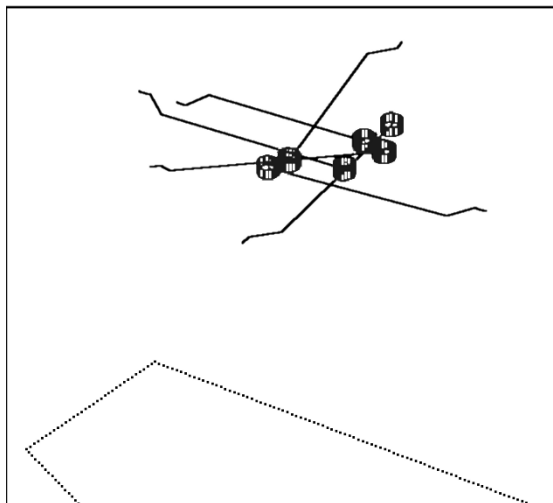


Figure 4.35. Resolution in 3-D.

maneuvers that allow the circles to possibly touch but always exquisitely avoiding overlapping.

As an interesting aside, notice from Fig. 4.32 that $\cap_{k=2}^6 I_{1k} = [4\mathbf{F}, 6\mathbf{F}]$. So any point on the interval $[4\mathbf{F}, 6\mathbf{F}]$ represents a path that eventually intersects *each one*

of the circles. This is an instance of the *one-shot problem* (OSP), which in general is NP-complete [133]. In 1989, another algorithm using particle fields on the *whole plane*, rather than just particle lines, was derived (US patent # 5,173,861). It also accomplished some resolutions with altitude changes; Fig. 4.35. It was reported²² that the US Air Force is testing the use of these algorithms for positioning several aircraft flying in formation [64] and for conflict detection and resolution [162].

Exercises

1. In Fig. 4.32 most of the associated times with the scrapes *increase*, since the point representing the scrape path is higher on the vertical line. Provide the conditions needed for these times to increase or decrease monotonically.
2. Provide the formulation for using several line fields $\ell(\theta)$ at different angles θ with the horizontal axis. Describe how you would efficiently construct the corresponding *conflict intervals* and their functional dependence on θ .
3. Define a particle field on the whole plane, as was done only for ℓ . Find the allowable particle regions (i.e., the set of particles satisfying the constraints) associated with:
 - (a) Maximum offset,
 - (b) Maximum angle,
 - (c) Conditions where the approach may fail.
4. Find the conditions for which the approach of “particle lines” fails.
5. Formulate the “one-shot problem” carefully. Describe conditions for which the conflict intervals can be used interactively to find one-shot solutions.

²²I am grateful to D. Sobiski for this information.

This page intentionally left blank

This page intentionally left blank

5

Planes, p -Flats, and Hyperplanes

5.1 *FT-1* Planes in \mathbb{R}^3

5.1.1 Vertical Line Representation

A hyperplane in \mathbb{R}^N can be translated to one that contains the origin, which is an $(N - 1)$ -dimensional linear subspace of \mathbb{R}^N . Since \mathbb{R}^{N-1} can be represented in \parallel -coords by $N - 1$ vertical lines and a polygonal line representing the origin, it is reasonable to expect a similar representation for hyperplanes in \mathbb{R}^N . An intuitive discussion for \mathbb{R}^3 clarifies matters. Consider a plane π as shown in Fig. 5.1, intersecting the x_1x_2 plane at the line y^1 and the x_2x_3 plane at the line y^2 with $A = y^1 \cap y^2$. The lines $y^i, i = 1, 2$, being lines in \mathbb{R}^3 , are represented by two points, $\bar{y}_{12}^i, \bar{y}_{23}^i$ each as shown. Next we construct a nonorthogonal coordinate system on π using the y^i as axes, consisting of the lines parallel to y^1 and the lines parallel to y^2 . Any point $P \in \pi$ can be specified as the intersection of two lines, one parallel to y^1 and the other parallel to y^2 . The family of lines parallel to y^1 is represented by a vertical line \bar{Y}_1 containing the point \bar{y}_{12}^1 . Similarly, the vertical line \bar{Y}_2 containing the point \bar{y}_{23}^2 represents the lines parallel to y^2 . Strictly speaking, the vertical lines \bar{Y}_i represent the *projections* on the x_1x_2 and x_2x_3 planes, respectively, of the two families of parallel lines. Therefore, the vertical lines represent two such families of parallel lines on *any* plane parallel to π . By choosing a point, say A , as the origin, we obtain a coordinate system specific to π . So the plane can be represented by two vertical lines and a polygonal line. Clearly, the same argument applies in any dimension, and therefore a hyperplane in \mathbb{R}^N can be represented by $N - 1$ vertical lines and a polygonal line representing one of its points [97]. Conversely, a set of coplanar points chosen on a grid such as

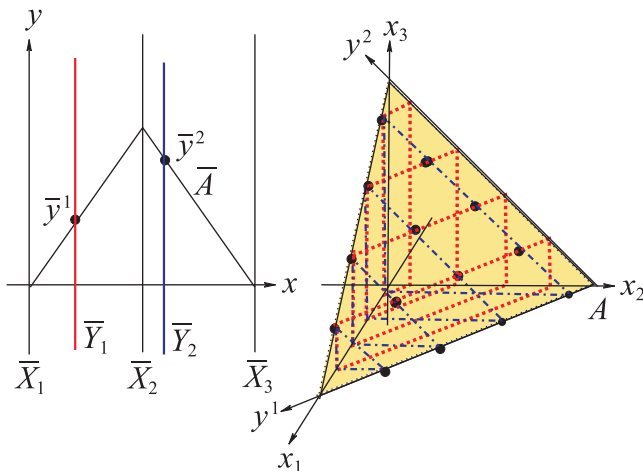


Figure 5.1. A plane π in \mathbb{R}^3 represented by two vertical lines and a polygonal line. This is a planar coordinate system with the polygonal line representing the origin.

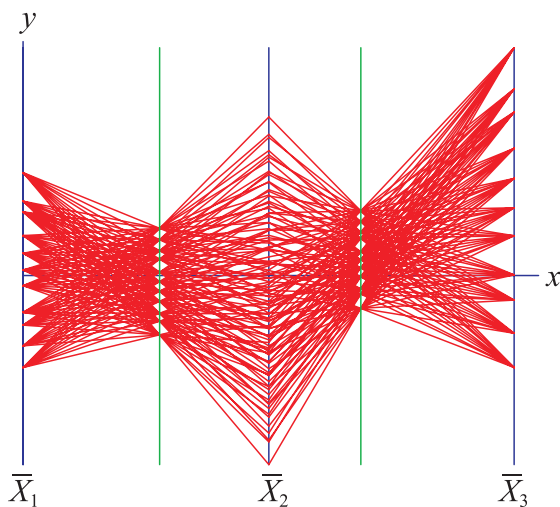


Figure 5.2. A set of coplanar points in \mathbb{R}^3 with the two vertical lines pattern.

the one formed by lines parallel to a coordinate system, as shown in Fig. 5.1, is represented by polygonal lines with a pattern specifying two vertical lines as shown in Fig. 5.2.

About the time these patterns were being discovered, word got around, and we were requested to visually explore for relations that may exist in a set of industrial

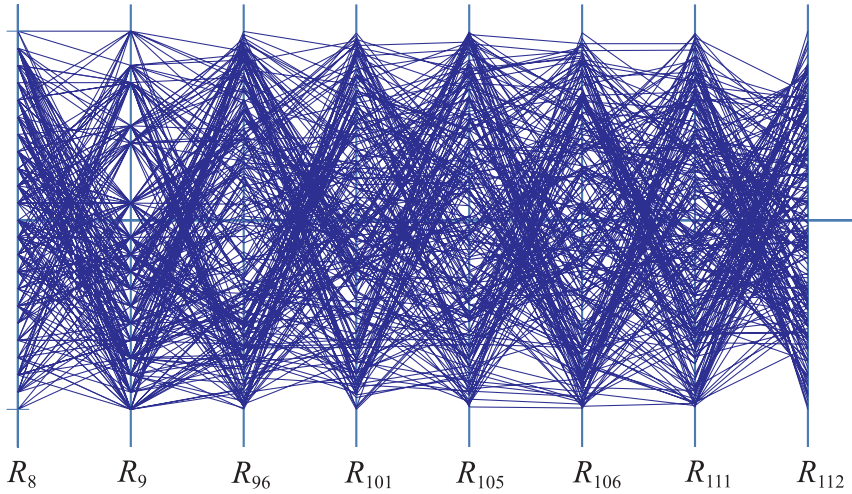


Figure 5.3. Industrial data with a “vertical line” pattern between R111 and R112 axes.

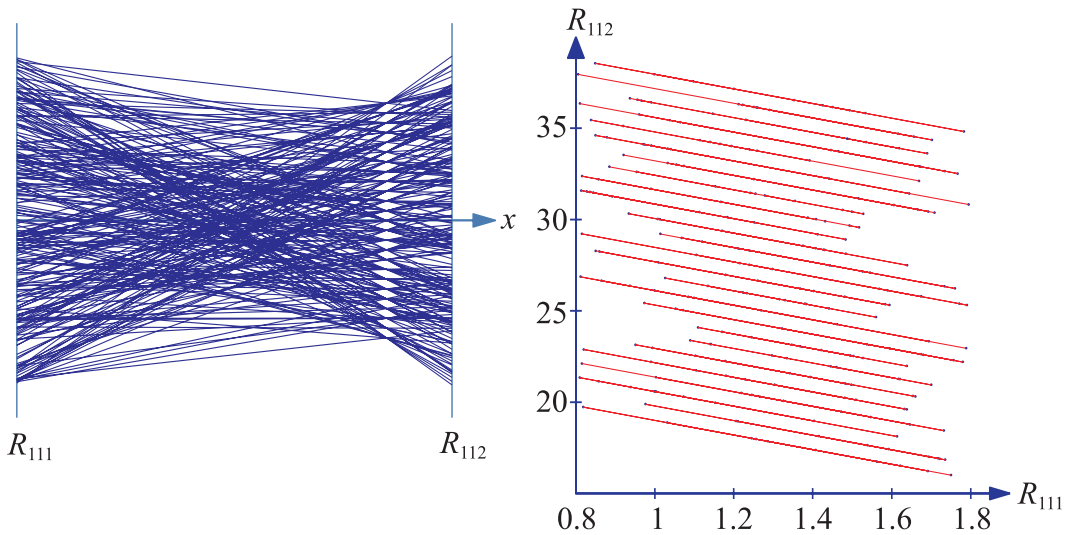


Figure 5.4. The portion between the R111 and R112 axes is magnified (left). This suggest that there is linear relation between R111, R112, and an unknown parameter, as is also indicated in the Cartesian coordinates plot on the right.

data consisting of several thousand records with eight variables. Part of this data set is plotted in parallel coordinates and shown in Fig. 5.3. The pattern between the R111 and R112 axes, which for clarity is magnified and shown in Fig. 5.4,

resembles one of vertical lines formed in Fig. 5.2. Yet for this pattern to represent a set of coplanar points, at least *two*, not one, vertical lines are needed, even for the plane with minimal dimensionality in R^3 . Still, the variables R111 and R112 are linearly interrelated, and there must be another variable, say X , also linearly related to R111 and R112, as is clear from Fig. 5.4 (see also the first exercise below). All this suggested that an important variable was not being measured. With these hints, a search was mounted and the missing variable was found. This, of course, was a stroke of good fortune and an instance of a favorite stochastic theorem, i.e., “You can’t be unlucky all the time.” Let us return from our digression. Before this is done, it is worth distinguishing the vertical line pattern between the R8 and R9 axes with that between the R111 and R112 axes. On the R8 axis and lower part of the R9 axis, the data are equally spaced, while this is not the case with the last two axes.

5.1.2 ** Planar Coordinates

The vertical lines representation of planes can also be obtained using the theory of envelopes,²³ starting with the description

$$\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0. \quad (5.1)$$

At first the envelope of the family of lines parallel to y^1 is found, as seen in Fig. 5.1. Two coordinates are needed to determine a point $P \in \pi$. One coordinate, say $x_3 = \alpha$, specifies a plane π' parallel to the x_1x_2 -plane. The line $\ell'(\alpha) = \pi \cap \pi'$ is in fact one of the lines parallel to y^1 ; see Fig. 5.7. It has the representation:

$$\bar{\ell}'(\alpha) : \begin{cases} \bar{\ell}'_{23}(\alpha) = (2, \alpha), \\ \bar{\ell}'_{12}(\alpha) = \left(\frac{c_2}{c_1+c_2}, \frac{c_0-c_3\alpha}{c_1+c_2} \right), \end{cases} \quad (5.2)$$

since

$$\ell'(\alpha) : \begin{cases} \ell'_{23}(\alpha) : x_3 = \alpha, \\ \ell'_{12}(\alpha) : c_1x_1 + c_2x_2 = c_0 - c_3\alpha, \end{cases} \quad (5.3)$$

where as usual, the distance between the parallel axes is one unit.

Returning to Fig. 5.2, each vertical line is formed by a set of points, and on the \bar{Y}_1 -axis these are the $\bar{\ell}'_{12}(\alpha)$. The polygonal lines intersecting at each such point represent points on the corresponding line $\ell'_{12}(\alpha)$. Together, these intersections provide one of the two vertical lines in the representation of π . As we know from

²³See the chapter on envelopes.

Chapter 3, for each α , $\bar{\ell}'_{12}(\alpha)$ is the *envelope* of a set of polygonal lines representing points on a line. As α varies, the $\bar{\ell}'_{12}(\alpha)$ form the vertical line \bar{Y}_1 , since, as one can see from (5.2), their x -coordinate is a constant (independent of α). Strictly speaking, the points $\bar{\ell}'_{23}(\alpha)$ also vary with α and fall on the \bar{X}_3 -axis. Since this would also occur for any other plane and the \bar{X}_3 -axis already exists, it does not contribute any information about the specific plane π , so typically it is not considered as part of the representation of a particular plane. This takes care of the lines on π parallel to the y^1 -axis. In exactly the same way it is found that the vertical line \bar{Y}_2 represents the lines parallel to the y^2 -axis.

What has effectively happened is that a coordinate system is constructed in the plane π consisting of two vertical lines and a polygonal line as an origin. The plane π is after all a 2-*fla*, i.e., it can be determined by three points. In general, an n -*fla* is a linear manifold (translate of a linear space) that is determined by $n + 1$ points. For an excellent short book on N -dimensional geometry see [163]. Now we would like to use this coordinate system to specify points on π . With reference to Fig. 5.5, consider a point $P \in \pi$. From its (x_1, x_2, x_3) coordinates we want to find its y^1 and y^2 coordinates or equivalently to express $P = \ell^1 \cap \ell^2$, where ℓ^1, ℓ^2 are lines on π and parallel to the y^1 - and y^2 -axes respectively. So we look at P_{12} , the projection of P on the x_1x_2 plane and construct the line through P_{12} parallel to y^1 . This line is represented by the point $\bar{\ell}^1_{12} = \bar{P}_{12} \cap \bar{Y}_1$ on the \bar{Y}_1 -axis. Similarly, we obtain $\bar{\ell}^2_{23} = \bar{P}_{23} \cap \bar{Y}_2$ on the \bar{Y}_2 -axis, where P_{23} is the projection of P on the x_2x_3 plane. Notice how careful tracking of the indices is already helpful in the construction. The pair $(\bar{\ell}^1_{12}, \bar{\ell}^2_{23})$ are the *planar coordinates* of \bar{P} in terms of the \bar{Y}_1 - and \bar{Y}_2 -axes. An interesting consequence is that a line $\ell \subset \pi$ can be represented by a single point $\bar{\eta}$ in, terms of the \bar{Y}_1, \bar{Y}_2 coordinate system. Of course, it is still represented by two points, the $\bar{\ell}$ s, in terms of the original $\bar{X}_i, i = 1, 2, 3$, Fig. 5.5.

Theorem 5.1.2. *For a line $\ell \subset \pi \subset \mathbb{R}^3$, the points $\bar{\ell}_{12}, \bar{\ell}_{23}$, and $\bar{\eta}$ are collinear.*

Proof. *Step 1.* With reference to the figure below, let $P_1, P_2 \in \ell$. The two lines AB and $A'B'$ joining the planar coordinates of \bar{P}_1 and \bar{P}_2 intersect at $\bar{\eta} = AB \cap A'B'$.

Step 2. The two triangles ABC and $A'B'C'$ formed between \bar{Y}_1 and \bar{Y}_2 are in perspective with respect to the ideal point in the vertical direction.

Step 3. The sides AB and CB are portions of \bar{P}_2 and correspond to $A'B'$ and $C'B'$, which are portions of \bar{P}_1 .

Step 4. By Step 3, $\bar{\ell}_{12} = AC \cap A'C'$ and $\bar{\ell}_{23} = BC \cap B'C'$.

Step 5. By Step 2 through Step 4 and Desargues's theorem, $\bar{\ell}_{12}, \bar{\ell}_{23}$, and $\bar{\eta}$ are all on the same line \bar{L} . ■

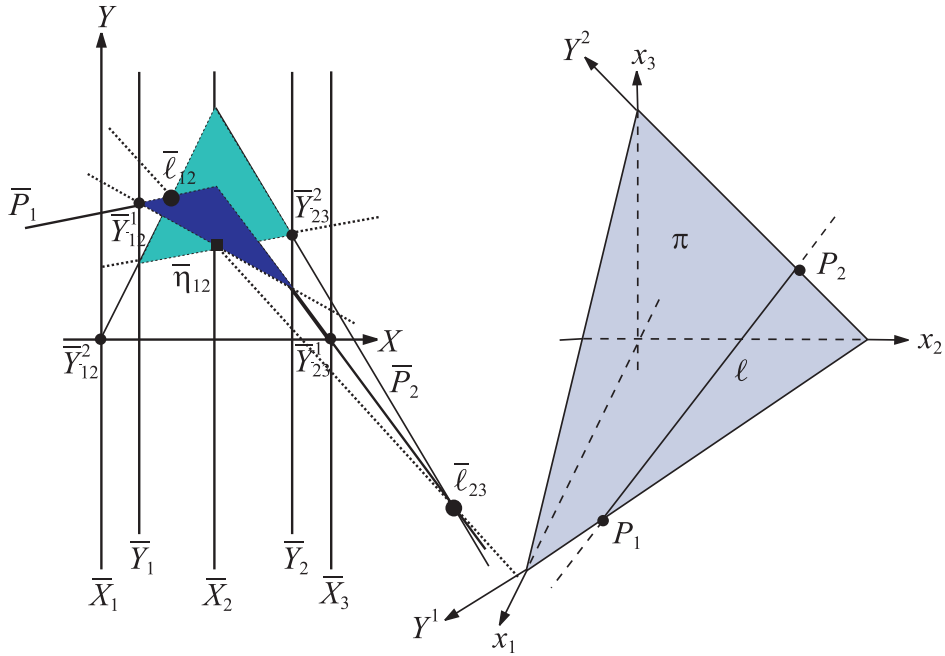


Figure 5.5. A line ℓ in a plane π is represented by one point $\bar{\eta}_{12}$. This is in terms of the planar coordinates \bar{Y}_1 and \bar{Y}_2 . The point $\bar{\eta}_{12}$ is collinear with the two points $\bar{\ell}_{12}$ and $\bar{\ell}_{23}$.

Of course, by the *3-point collinearity property* of Chapter 4, the point $\bar{\ell}_{13}$ is also on \bar{L} .

Corollary 5.1.3. *The rotation of a plane in \mathbb{R}^3 about a line corresponds to a translation of a point on a line.*

Proof. With reference to the Fig. 5.6, let $\ell \subset \pi \subset \mathbb{R}^3$. Rotate π about ℓ to a new position π^* . Let $\bar{\ell}_{12}, \bar{\ell}_{23}$ represent ℓ in the $\bar{X}_1, \bar{X}_2, \bar{X}_3$ coordinates and $\bar{\eta}, \bar{\eta}^*$ in the \bar{Y}_1, \bar{Y}_2 and \bar{Y}_1^*, \bar{Y}_2^* coordinate systems respectively. By the theorem, $\bar{\ell}_{12}, \bar{\ell}_{23}$, and $\bar{\eta}$ are on a line \bar{L} . Also, $\bar{\ell}_{12}, \bar{\ell}_{23}$, and $\bar{\eta}^*$ are on the line \bar{L} . That is, the rotation of π about ℓ corresponds in $\|\text{-coords}$ to a translation of the point $\bar{\eta}$ on \bar{L} . ■

This is the 3-D analogue of (*rotation of a line around a point*) \rightarrow (*translation of a point along a line*) in 2-D and corresponds to the *point \leftrightarrow plane* duality in \mathbb{P}^3 . Let us explore the computation involved, where ℓ is given by

$$\ell : \begin{cases} \ell_{12} : & x_2 = m_2 x_1 + b_2, \\ \ell_{23} : & x_3 = m_3 x_2 + b_3. \end{cases} \quad (5.4)$$

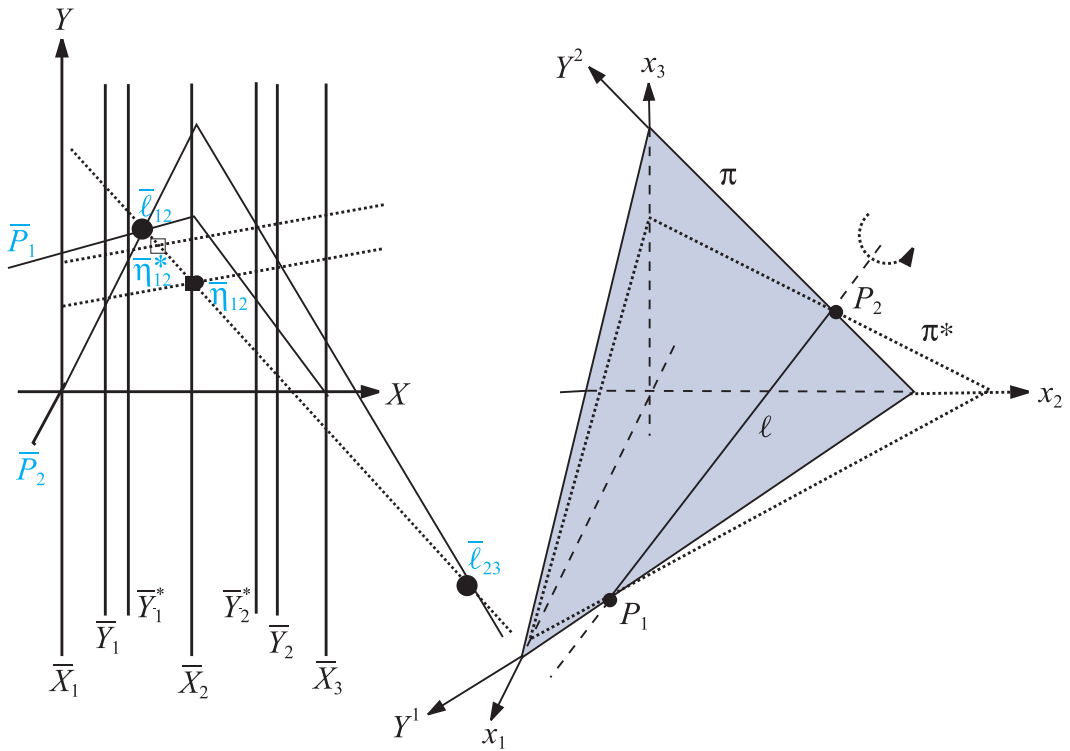


Figure 5.6. Rotation of a plane about a line \leftrightarrow translation of a point along a line.

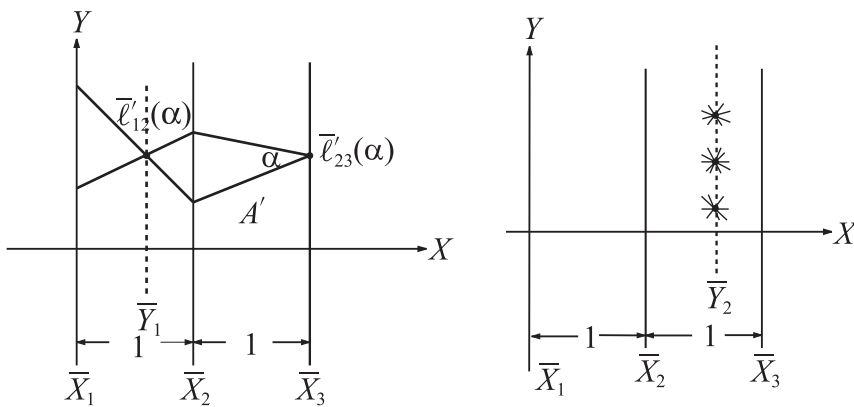


Figure 5.7. The construction of \bar{Y}_1 (left), the first, and \bar{Y}_2 (right), the second, vertical axes. Start with a line on π parallel to the y^1 and y^2 axes respectively.

There exists a one-parameter, say k , family²⁴ of planes containing ℓ . In fact, the equation of any one of the planes in the family is given by

$$\pi : (x_3 - m_3x_2 - b_3) + k(x_2 - m_2x_1 - b_2) = 0. \quad (5.5)$$

It can be verified that in Cartesian coordinates,

$$\bar{\eta}_{12} = \left(\frac{m_3^2 - 2m_3 - k^2}{m_3^2 - m_3 + k^2(m_2 - 1)}, -\frac{b_2k^2 + m_3b_3}{m_3^2 - m_3 + k^2(m_2 - 1)} \right). \quad (5.6)$$

Every value of k corresponds to a position (of the rotated plane) π and, in turn, about a position (of the translated point). η along the line formed by the points $\bar{\ell}_{12}$ and $\bar{\ell}_{23}$.

The generalization for \mathbb{R}^N is straightforward and is not covered here (see [97]).

Fun and Games with the ILM

Open the ILM2. On the right is a plane π containing a line ℓ . The plane is represented by two vertical lines \bar{Y}_1, \bar{Y}_2 on the left. The line ℓ is specified by two points P_1, P_2 (right) and represented by $\bar{\ell}_{12}, \bar{\ell}_{23}$ on the left. The point $\bar{\eta}_{12}$ represents the line ℓ with respect to the planar axes \bar{Y}_1, \bar{Y}_2 . All three points $\bar{\ell}_{12}, \bar{\ell}_{23}, \bar{\eta}_{12}$ are on a line \bar{L} .

For some experimentation: slide the point $\bar{\pi}_{123}$ along \bar{L} ; notice the translation of $\bar{\eta}_{12}$ and corresponding rotation of the plane. Change the line and/or plane and repeat. Click other buttons and experiment.

Exercises

1. The pattern between the R111 and R112 axes in Figures 5.3 and 5.4 raises a question. What if a different permutation of the axes were chosen? Whereas the coplanarity information would still be preserved, the “vertical line pattern” would not be seen. Was it fortuitous that the “right” axes permutation was chosen to plot the data? It is not difficult to show that for N variables, there are $O(N/2)$ “cleverly” chosen permutations that provide all possible *adjacent*

²⁴In the language of projective geometry this is called a *pencil* of planes on the line.

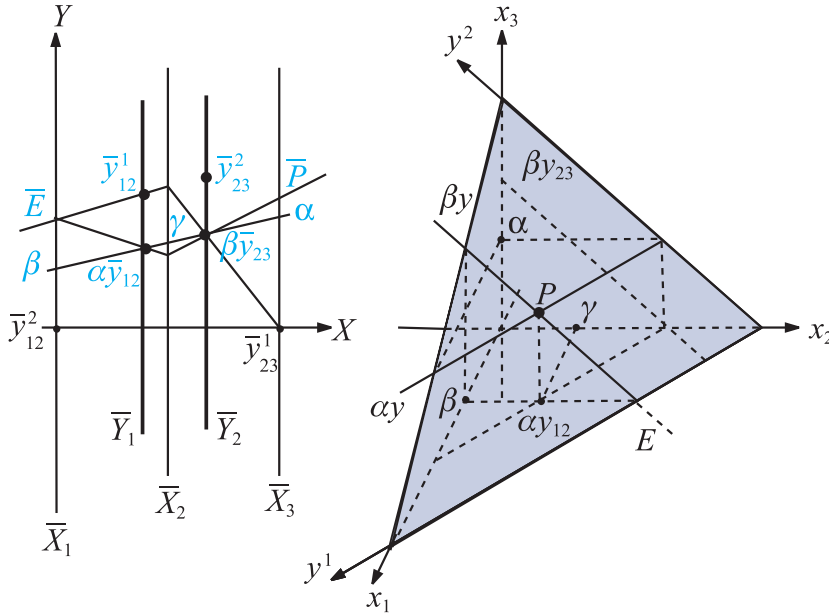


Figure 5.8. Constructing a point in a plane.

pairs of axes. For example, for $N = 6$, three such permutations are (just showing the subscripts of the \bar{X}_i) 126354, 231465, 342516 [184]. Returning to the example, there are four such permutations, one of which has the coplanarity pattern.

- (a) Prove this permutation result. Hint: construct an undirected graph whose vertices are variables, and edges between a pair of vertices designate that the corresponding vertices are adjacent in a particular permutation.
 - (b) Find an algorithm for computing these special $O(N/2)$ permutations.
 - (c) Generalize this result for adjacent *triples*.
 - (d) Generalize this result for adjacent *p-tuples*.
2. Given three noncollinear points in \mathbb{R}^3 , find an algorithm for constructing in $\|\text{-coords}$ a set of planar coordinates for the plane determined by the three points.
 3. Construct a set of planar coordinates for a plane π perpendicular to one of the principal 2-planes (i.e., π being parallel to one of the orthogonal axes).
 4. Given a set of planar coordinates for a plane π , find an algorithm for constructing a plane π' perpendicular to it.
 5. In $\|\text{-coords}$ find an algorithm for determining the intersection of two planes.

6. Given a plane π and a line ℓ , provide an algorithm that in \parallel -coords determines whether $\ell \subset \pi$.
7. Given a plane π and a point P , provide an algorithm for determining whether P is below, on, or above the plane. For the notion of above/below, consider π partitioning \mathbb{R}^3 in to two half spaces that need to be distinguished. All points in one half space are “above,” etc.
8. Provide an algorithm in \parallel -coords for intersecting a line with a plane.
9. Verify,
 - (a) (5.5),
 - (b) (5.6).

♣ FT-1e

5.2 **♣ FT-2** Representation by Indexed Points

The representation of a plane in terms of vertical lines is basically the representation of a specific coordinate system in the plane. With this representation, the coplanarity of a set of points can be checked visually only if the points are on a rectangular grid as in Fig. 5.2. In Fig. 5.9 the polygonal lines shown on the $\bar{X}_1, \bar{X}_2, \bar{X}_3$ -axes represent

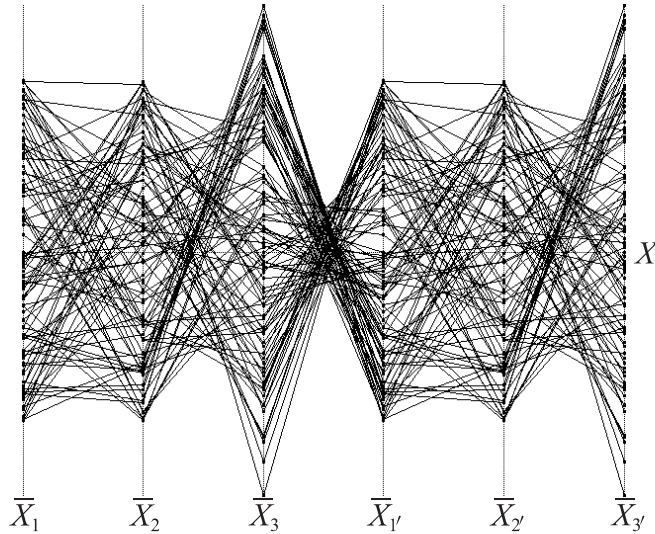


Figure 5.9. Randomly selected coplanar points in \mathbb{R}^3 represented by the polygonal lines on the first three axes.

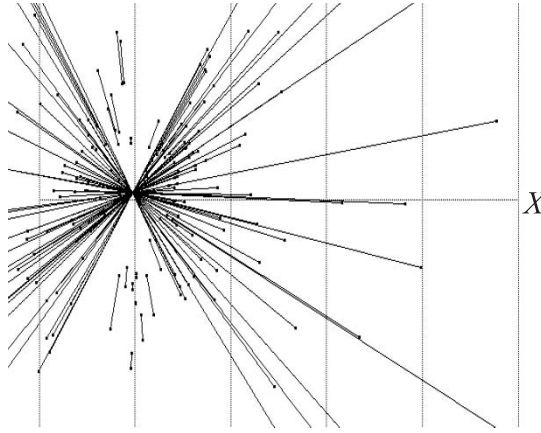


Figure 5.10. Coplanarity!

A pencil of lines on a point is formed by joining the pairs of points representing lines on a plane.

randomly sampled points in a plane π in \mathbb{R}^3 . There is no discernible pattern. Is there a \parallel -coords “pattern” associated with coplanarity of randomly selected points?

A new approach is called for [47]. Describing even a simple 3-dimensional object such as a room in terms of its points is just not intuitive. The description of it in terms of cross-sections including the boundary planes helps us visualize the object. Pursuing the analogy, it is proposed to describe p -dimensional objects in terms of their $(p - 1)$ -dimensional subsets rather than their points. So the representation of a plane π is attempted in terms of the *lines* it contains. From each pair of polygonal lines ℓ shown in Fig. 5.9, the representation of the corresponding line $\ell \subset \pi$ is constructed. The result in Fig. 5.10 is stunning: the lines joining the pairs of points $\bar{\ell}_{12}, \bar{\ell}_{23}$ in turn determine a pencil of lines on a point, and this is characteristic of coplanarity. Let us explore it.

5.2.1 The family of “Superplanes” \mathcal{E}

Behind the striking pattern in Fig. 5.10 lurks a very special subspace of \mathbb{R}^N . Until now, the \parallel -coords axes were taken equidistant, with the y - and \bar{X}_1 -axes coincident. This was a matter of convenience, which has served us well until now. The time has come to look at the general setting shown in Fig. 5.11, the position of the \bar{X}_i specified by the directed (i.e., signed) distance d_i , with the stipulation that for some $i \neq j$, $d_i \neq d_j$. We consider the set of points $P \in \mathbb{R}^N$ whose representation in \parallel -coords collapses to a straight line. That is, $\bar{P} : y = mx + b$, and for a specific choice of (m, b) , the corresponding point is

$$P = (md_1 + b, \dots, md_N + b) = m(d_1, d_2, \dots, d_N) + b(1, \dots, 1). \quad (5.7)$$

For $m, b \in \mathbb{R}$ this collection of points forms the subspace π^s of \mathbb{R}^N spanned by the two N -tuples (d_1, d_2, \dots, d_N) , $(1, \dots, 1)$. It contains the points $(0, 0, \dots, 0)$ and $(1, 1, \dots, 1)$ and the complete line u on these points. For future reference, the unit vector on u anchored at the origin and pointing towards $(1, 1, \dots, 1)$ is denoted by \mathbf{u} . The two-dimensional subspaces on the line u are called *superplanes* and play a crucial role in our development. The family of superplanes generated by all possible axes spanning (i.e., all (d_1, d_2, \dots, d_N) , $(1, \dots, 1)$) is denoted by \mathcal{E} , after J. Eickenmeyer, who first discovered them. The situation is especially interesting in \mathbb{R}^3 where the superplanes are

$$\pi^s = \{m(d_1, d_2, d_3) + b(1, 1, 1) | m, b \in \mathbb{R}\},$$

and whose equation

$$\pi^s : (d_3 - d_2)x_1 + (d_1 - d_3)x_2 + (d_2 - d_1)x_3 = 0 \quad (5.8)$$

can be obtained by using numerically convenient values (i.e., $m = b = 0$, $m = 0, b = 1$, $m = 1, b = 0$). It describes the *pencil* of planes, on the line u each specified by the spacing of the axes. As for (5.5), it can be rewritten as a one-parameter family:

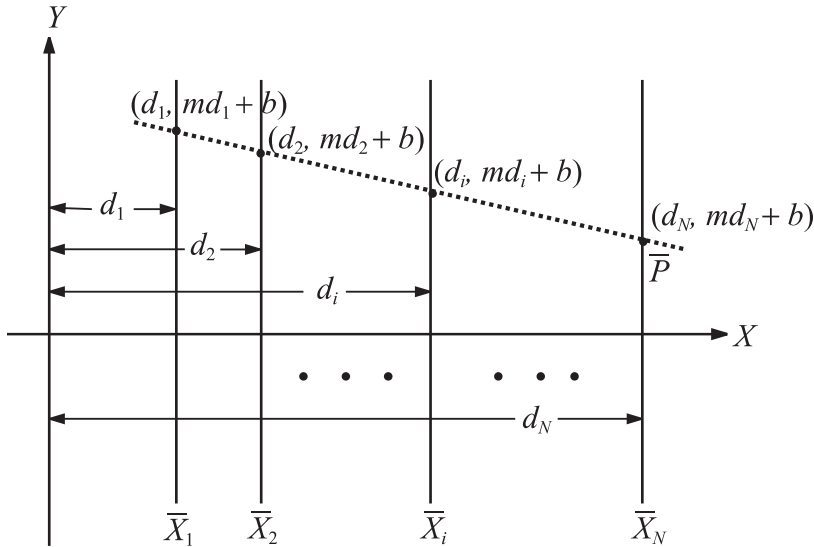


Figure 5.11. Points in \mathbb{R}^N represented by straight (rather than polygonal) lines.

$$\pi^s : (x_3 - x_2) + k(x_2 - x_1) = 0, k = \frac{d_2 - d_3}{d_2 - d_1}, \quad (5.9)$$

with the ratio k determining the particular plane. With the y - and \bar{X}_1 -axes coincident and the standard axes spacing $d_1 = 0, d_2 = 1, d_3 = 2$, the superplane (sometimes referred to as the “first superplane”) is

$$\pi_1^s : x_1 - 2x_2 + x_3 = 0. \quad (5.10)$$

The subscript 1 distinguishes it from the others formed with different spacing of the axes. The important property from Chapter 4 can now be restated in terms of π_1^s .

Theorem 5.2.1 (3-point collinearity) *For any line $\ell \subset \mathbb{R}^N$, the points $\bar{\ell}_{ij}, \bar{\ell}_{jk}, \bar{\ell}_{ik}$, where the i, j, k are distinct integers $\in [1, 2, \dots, N]$, are on a line \bar{L} , where $L = \ell \cap \pi_1^s$.*

This is the backbone of the recursive (in the dimension) construction algorithm that follows.

5.2.2 The Triply Indexed Points

Let us now review the construction leading to the single intersection in Fig. 5.10. For any line $\ell \subset \pi$, the three points $\bar{\ell}_{12}, \bar{\ell}_{23}, \bar{\ell}_{13}$ are:

- on a line \bar{L} by the 3-point-collinearity property, and conversely, since \bar{L} lies on the three points, the point L is on ℓ .
- Further, since \bar{L} is a straight line, L must also be a point of the “superplane” π_1^s .

Therefore

$$\ell \cap \pi_1^s = L. \quad (5.11)$$

This is true for *every* line $\ell \subset \pi$. Specifically, for a line

$$\ell' \subset \pi, \quad \exists L' \in \ell' \quad \ni \ell' \cap \pi^s = L'.$$

Now L and L' specify a line $\ell_\pi \subset \pi_1^s$ represented by a single point $\bar{\ell}_\pi$. Altogether,

$$\left. \begin{array}{ll} L, L' \in \pi & \Rightarrow \ell_\pi \subset \pi \\ L, L' \in \pi_1^s & \Rightarrow \ell_\pi \subset \pi_1^s \end{array} \right\} \Rightarrow \text{for } \pi \neq \pi_1^s, \pi \cap \pi_1^s = \ell_\pi,$$

showing that the point where all the lines intersect in Fig. 5.10 is $\bar{\ell}_\pi$. For reasons that are clarified shortly, we refer to $\bar{\ell}_\pi$ by $\bar{\pi}_{123}$.

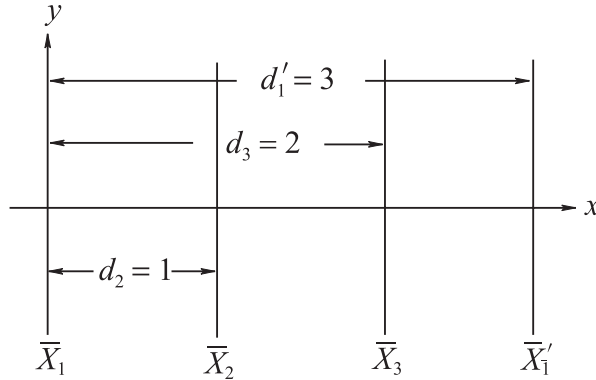


Figure 5.12. The axes spacing for the second superplane π_1^s .

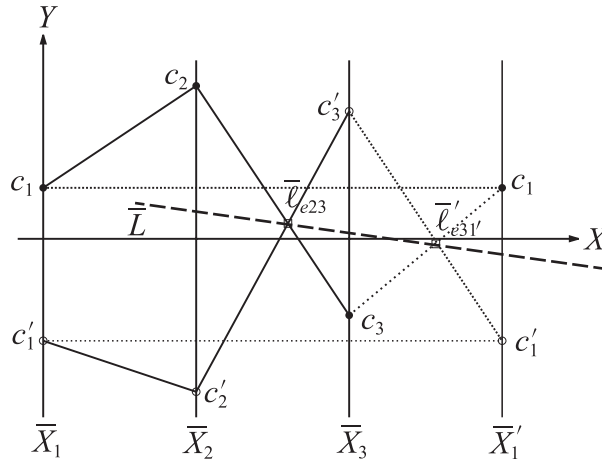


Figure 5.13. Transferring the values from the \bar{X}_1 -axis to the $\bar{X}_{1'}$ -axis.

A plane $\pi \subset \mathbb{R}^3$ is determined by two intersecting lines. It is advantageous to use two lines belonging to “superplanes,” since their representation requires only one point. One such line is $\ell_\pi = \pi \cap \pi_1^s$. To determine the second line we revisit (5.8) and choose another convenient spacing of the axes. As indicated in Fig. 5.9, the \bar{X}_1 -axis is translated (recall that this corresponds to a rotation) to $\bar{X}_{1'}$, one unit to the right of \bar{X}_3 , as shown in Fig. 5.12 with $d_2 = 1$, $d_3 = 2$, $d_1' = 3$. The new superplane is given by

$$\pi_{1'}^s : x_1 + x_2 - 2x_3 = 0. \quad (5.12)$$

Transferring the x_1 coordinate of each point P to the \bar{X}'_1 -axis as shown in Fig. 5.13 and repeating the construction, we obtain the second point, denoted by $\bar{\pi}_{231}'$ and shown in Fig. 5.15. The corresponding line is none other than the intersection $\ell'_\pi = \pi \cap \pi_1^s$. What about the line \bar{H} joining the points $\bar{\pi}_{123}$ and $\bar{\pi}_{231}'$? Clearly, $H \in \ell_\pi$, since $\bar{\pi}_{123}$ represents ℓ_π and similarly $H \in \ell'_\pi$. Further, since \bar{H} is a line in the $\bar{X}_1, \bar{X}_2, \bar{X}_3$ coordinate system and also in the $\bar{X}_2, \bar{X}_3, \bar{X}'_1$ system,

$$H = \ell_\pi \cap \ell'_\pi = \pi_1^s \cap \pi_1^{s'} \cap \pi, \quad (5.13)$$

as illustrated in Fig. 5.14. So the intersecting lines ℓ_π, ℓ'_π determine the plane π , and therefore the plane can be represented by the *two points* $\bar{\pi}_{123}$ and $\bar{\pi}_{231}'$. But what is going on? In Chapter 4 we saw that a *line* in \mathbb{R}^3 is also represented by two points. To distinguish the two points representing a plane, *three* indices are attached, while only two indices are used for the point representation of a line. The deeper reason is, of course, that the linear relations specifying lines involve (or can

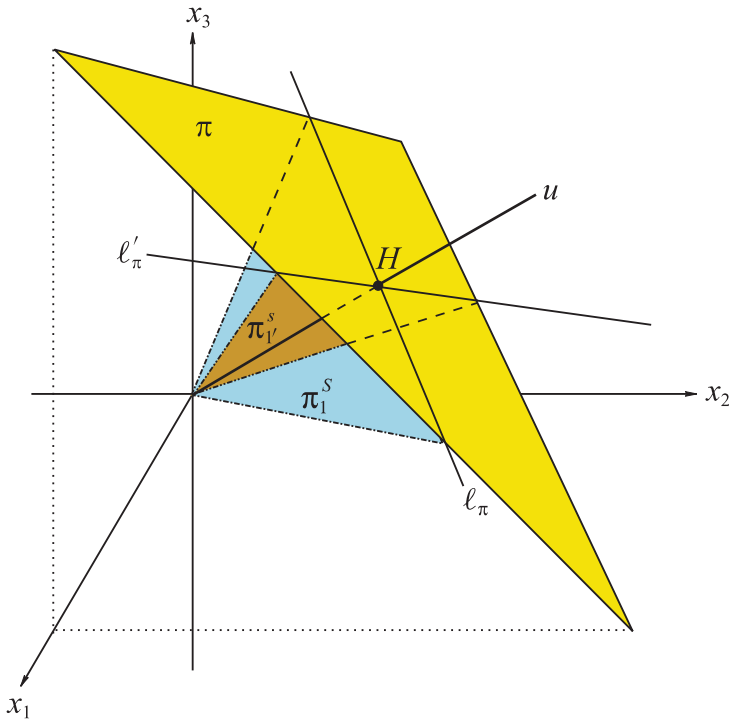


Figure 5.14. The intersections of a plane π with the two superplanes π_1^s and $\pi_1^{s'}$. These are the two lines ℓ_π, ℓ'_π that specify the plane and provide its representation.

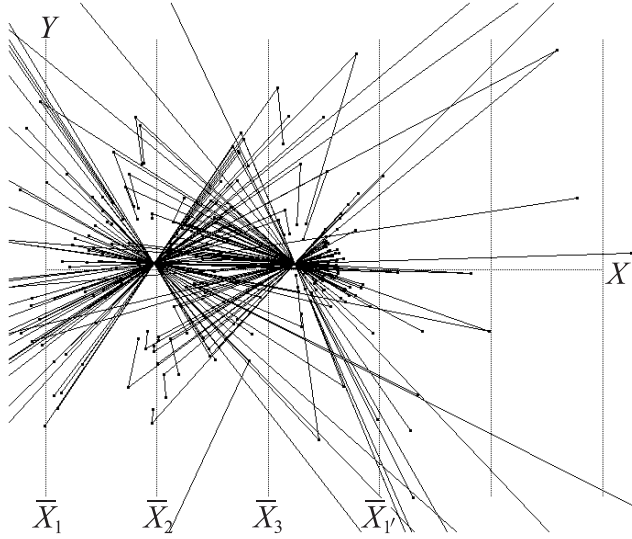


Figure 5.15. The plane π represented by two points.

be reduced to) two variables, whereas those involving planes necessarily involve three variables. Specifically, for the plane,

$$\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0, \quad (5.14)$$

$$\ell_\pi = \pi \cap \pi_1^s : \begin{cases} \ell_{\pi_{12}} : & x_2 = -\frac{c_1 - c_3}{c_2 + 2c_3}x_1 + \frac{c_0}{c_2 + 2c_3}, \\ \ell_{\pi_{23}} : & x_3 = -\frac{2c_1 + c_2}{c_3 - c_1}x_2 + \frac{c_0}{c_3 - c_1}. \end{cases} \quad (5.15)$$

Therefore, in homogeneous coordinates, recalling that 1, the distance of \bar{X}_2 from the y -axis, must be added to the first coordinate of $\bar{\ell}'_{\pi_{23}}$,

$$\bar{\pi}_{123} = \bar{\ell}_{\pi_{12}} = \bar{\ell}_{\pi_{23}} = (c_2 + 2c_3, c_0, c_1 + c_2 + c_3). \quad (5.16)$$

Continuing,

$$\ell'_\pi = \pi \cap \pi_{1'}^s : \begin{cases} \ell'_{\pi_{12}} : & x_2 = -\frac{2c_1 + c_3}{2c_2 + c_3}x_1 + \frac{2c_0}{2c_2 + c_3}, \\ \ell'_{\pi_{23}} : & x_3 = -\frac{c_2 - c_1}{2c_1 + c_3}x_2 + \frac{c_0}{2c_1 + c_3}, \end{cases} \quad (5.17)$$

and from the second equation it is immediate that

$$\bar{\ell}'_{\pi_{23}} = (3c_1 + c_2 + 2c_3, c_0, c_1 + c_2 + c_3). \quad (5.18)$$

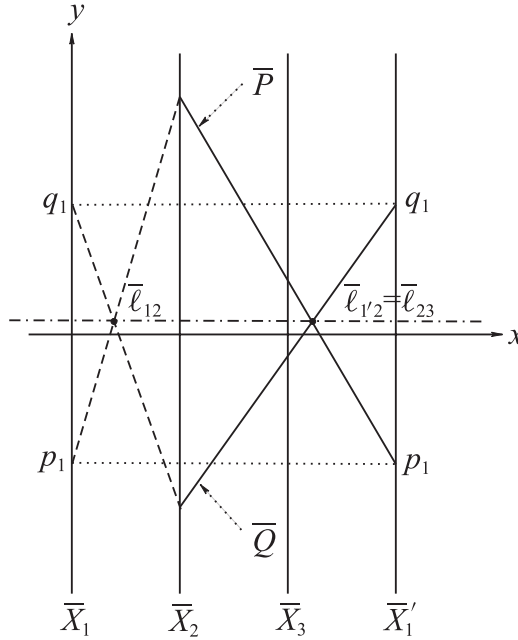


Figure 5.16. The location of the $\bar{\ell}_{12}$ and $\bar{\ell}_{1'2}$ points.

Since ℓ'_π is a line in the superplane $\pi^s_{1'}$, it must be that $\bar{\ell}'_{\pi_{12}} = \bar{\ell}'_{\pi_{23}}$. Yet a direct computation from the first equation of (5.17) yields

$$\bar{\ell}'_{\pi_{12}} = (2c_2 + c_3, c_0, 2(c_1 + c_2 + c_3)). \quad (5.19)$$

What is going on? Figure 5.16 explains the “riddle,” reminding us that $\pi^s_{1'}$ is a superplane in the $\bar{X}_2, \bar{X}_3, \bar{X}_{1'}$ coordinate system, where its points such as P, Q are represented by straight lines intersecting at $\bar{\ell}'_{\pi_{1'2}} = \bar{\ell}'_{\pi_{23}}$. To obtain $\bar{\ell}'_{\pi_{12}}$, the x_1 values must be transferred to the \bar{X}_1 -axis and then the corresponding *polygonal lines* constructed to obtain their intersection at $\bar{\ell}'_{\pi_{12}} \neq \bar{\ell}'_{\pi_{1'2}}$. Recall from Chapter 3 the line $\ell \rightarrow \bar{\ell}$ point correspondence

$$\ell : x_2 = mx_1 + b \rightarrow \bar{\ell} : \left(\frac{d}{1-m}, \frac{b}{1-m} \right), \quad m \neq 1, \quad (5.20)$$

where d is the *directed* interaxis distance. The distances from $\bar{X}_{1'}$ to \bar{X}_2 and the y -axis are two and three units respectively. Then together with the first equation in (5.17),

$$\bar{\ell}'_{\pi_{1'2}} = \left(\frac{-2}{1 + \frac{2c_1+c_3}{2c_2+c_3}} + 3, \frac{c_0}{1 + \frac{2c_1+c_3}{2c_2+c_3}} \right), \quad (5.21)$$

which in homogeneous coordinates matches (5.19), and in analogy to (5.16), we record²⁵ the result as

$$\bar{\pi}_{231'} = \bar{\ell}'_{\pi_{1'2}} = \bar{\ell}'_{\pi_{23}} = (3c_1 + c_2 + 2c_3, c_0, c_1 + c_2 + c_3). \quad (5.22)$$

To simplify the notation, we also write $\bar{\pi}_{1'} = \bar{\pi}_{1'23} = \bar{\pi}_{231'}$ and $\bar{\pi}_{0'} = \bar{\pi}_{123}$. The coordinates of the two points $\bar{\pi}_{0'}$ and $\bar{\pi}_{1'}$ contain three independent parameters and suffice to determine the coefficients of π . Let $S = c_1 + c_2 + c_3$ and denote by $x_{0'}, x_{1'}$ the x Cartesian coordinates of $\bar{\pi}_{123}$ and $\bar{\pi}_{1'23}$ respectively when $S \neq 0$. Then

$$x_{1'} - x_{0'} = 3 \frac{c_1}{S} = 3c'_1, \quad (5.23)$$

where the $c'_i = c_i/S, i = 0, 1, 2, 3$, are the *normalized* coefficients. Exploring this further, the \bar{X}_2 -axis is translated to the right by three units, the construction is repeated, and a third point is obtained:

$$\bar{\pi}_{31'2'} = \bar{\pi}_{1'2'3} = (3c_1 + 4c_2 + 2c_3, c_0, c_1 + c_2 + c_3), \quad (5.24)$$

which is also denoted by $\bar{\pi}_{2'}$ and its x coordinate by $x_{2'}$. Then

$$x_{2'} - x_{1'} = 3c'_2. \quad (5.25)$$

Finally, by translating the \bar{X}_3 -axis three units to the right and repeating the construction as shown in Fig. 5.17, a fourth point is obtained,

$$\bar{\pi}_{1'2'3'} = \bar{\pi}_{3'} = (3c_1 + 4c_2 + 5c_3, c_0, c_1 + c_2 + c_3), \quad (5.26)$$

and for $x_{3'}$, its x coordinate

$$x_{3'} - x_{2'} = 3c'_3. \quad (5.27)$$

Clearly, the third and fourth points are dependent on the first two, and by an easy calculation it is found that

$$x_{2'} = 6 - (x_{0'} + x_{1'}), \quad x_{3'} = 3 + x_{0'}. \quad (5.28)$$

So with the normalization $c'_1 + c'_2 + c'_3 = 1$, the distance between adjacent (according to the indexing) points of the plane's representation is proportional to the corresponding coefficient. The proportionality constant equals the dimensionality of the space (see Exercise 5). Their ordinate is the constant c'_0 , so the equation of the plane

²⁵Another detailed example is computed later for the principal 2-planes.

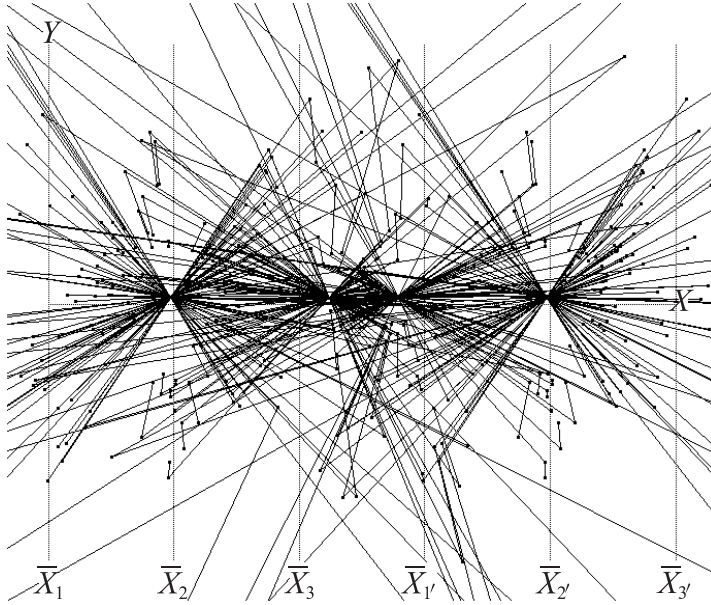


Figure 5.17. The plane π intersected with four superplanes. Each point represents one of the resulting lines.

can effectively be read from the picture as shown in Fig. 5.18. Occasionally, we say that the distance between adjacent indexed points *is* the corresponding coefficient when the dimensionality and hence the proportionality constant are clear. When the coefficients' sum is zero, the plane is an *sp* represented by ideal points. All the superplanes are on the line u , so what is the angle between the four different superplanes we generated? To answer this, it is convenient to introduce and work with vector notation. Letting $\mathbf{x} = (x_1, x_2, x_3)$, the coefficients of a plane π^j are $\mathbf{c}^j = (c_1^j, c_2^j, c_3^j)$, so the plane's equation is $\pi^j : \mathbf{c}^j \cdot \mathbf{x} = c_0^j$, where “ \cdot ” stands for the inner (or dot) product. From analytic geometry, the angle ϕ between two planes π^1, π^2 is found by

$$\cos \phi = \pm \frac{\mathbf{c}^1 \cdot \mathbf{c}^2}{[(\mathbf{c}^1 \cdot \mathbf{c}^1)(\mathbf{c}^2 \cdot \mathbf{c}^2)]^{\frac{1}{2}}}. \quad (5.29)$$

Fixing the angle by adopting the $+$ sign above, we find that π_1^s is obtained from π_1^s by a clockwise rotation of 120° about u . Checking for the third superplane

$$\pi_{1'2'3}^s : -2x_1 + x_2 + x_3 = 0 \quad (5.30)$$

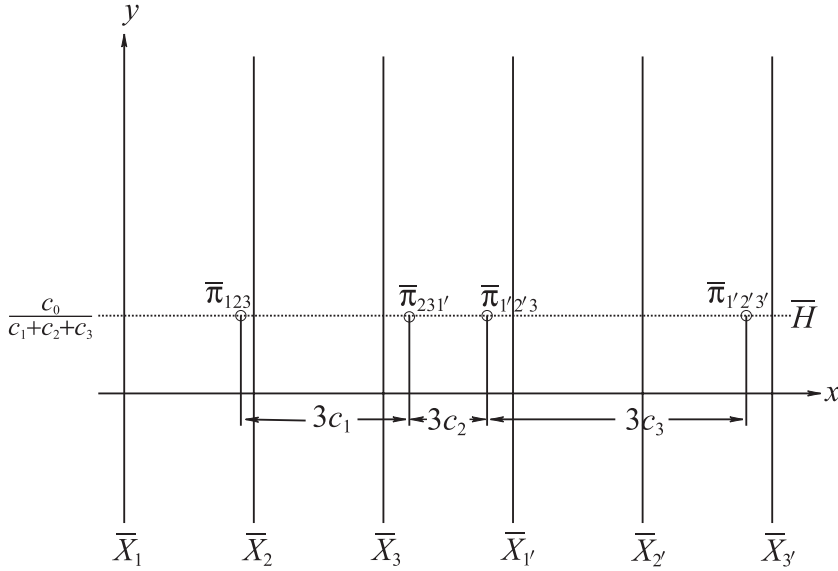


Figure 5.18. The distances between adjacent points are proportional to the coefficients.

For $\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0$ with the normalization $c_1 + c_2 + c_3 = 1$, the proportionality constant is the dimensionality of the space. The plane's equation can be read from the picture!

resulting from the translation of the \bar{X}_2 -axis, the dihedral angle with $\pi_{1'}^s$ is also 120° . Not surprisingly, the fourth superplane $\pi_{1'2'3'}^s$, generated by the translation of the \bar{X}_3 -axis, coincides with π_1^s .

To clarify, the rotations about u and the superplanes they generate do not affect the plane π , which remains *stationary*, intersecting u at H given in (5.13) and shown in Fig. 5.14. This is the reason that all points $\bar{\pi}_{123}, \bar{\pi}_{1'23}, \bar{\pi}_{1'2'3}, \bar{\pi}_{1'2'3'}$ have the same y -coordinate $c_0/(c_1 + c_2 + c_3)$ on the horizontal line \bar{H} . This useful property is also true for the points representing lines and is put to good use in the construction algorithms discussed next. For consistency with this chapter's format the equations of a line ℓ are rewritten as

$$\ell : \begin{cases} \ell_{1,2} : c_{11}x_1 + c_{12}x_2 &= c_{10}, \\ \ell_{2,3} : c_{22}x_2 + c_{23}x_3 &= c_{20}, \end{cases} \quad (5.31)$$

emphasizing that each of the $\ell_{1,2}, \ell_{2,3}$ is a *projecting plane* in \mathbb{R}^3 orthogonal to the x_1x_2, x_2x_3 planes respectively with $c_{13} = c_{21} = 0$. It is clear, then, that the points $\bar{\ell}_{12}, \bar{\ell}_{1'2}, \bar{\ell}_{1'2'}$ all have the same y -coordinate and are on a horizontal line; let us call it **12**. Similarly, the triple $\bar{\ell}_{23}, \bar{\ell}_{2'3}, \bar{\ell}_{2'3'}$ are on a horizontal line **23**, and

$\bar{\ell}_{13}, \bar{\ell}_{1'3}, \bar{\ell}_{1'3'}$ are on another horizontal line **13**. All this is seen in Fig. 5.25 of the next section.

Exercises

1. Derive the equation equivalent to (5.8) for \mathbb{R}^N and state its properties.
2. What planes are represented when the points $\bar{\pi}_{123}$ and $\bar{\pi}_{231'}$ are ideal?
3. How are the planes in class \mathcal{E} represented?
4. Is the line intersection shown in Fig. 5.10 a necessary and sufficient condition for coplanarity? Prove or give a counterexample.
5. Show that in \mathbb{R}^N , the distance between adjacent indexed points is N times the corresponding coefficient, as in \mathbb{R}^3 it is 3; equations (5.23), (5.25), and (5.27).
6. (a) Find the equations of the three superplanes containing the x_1 -, x_2 - and x_3 -axes respectively.
(b) What are the three dihedral angles between pairs of these superplanes?
7. Provide an affine transformation resulting in the four points $\bar{\pi}_{123}, \bar{\pi}_{231'}, \bar{\pi}_{31'2'}, \bar{\pi}_{1'2'3'}$ being collinear and with distance c_i between pairs rather than $3c_i$.
8. The vertical line representation of planes actually requires four rather than the two vertical lines shown. Explore the case in which none of the four vertical lines coincides with the coordinate axis \bar{X}_i .
9. Provide algorithms based on the indexed-points representation for finding the intersection between:
 - (a) a pair of 2-flats, and
 - (b) a line and a 2-flat.
 Delimit carefully the special cases in which the algorithms fail.
10. Given the representation of a plane in terms of vertical lines,
 - (a) how would one check whether the vertical lines represent two *orthogonal* families of parallel lines?
 - (b) How could the vertical axes be transformed to vertical axes representing *orthogonal* families of parallel lines?
11. Construct a vertical line representation \bar{Y}_i for a plane $\pi \subset \mathbb{R}^3$ with $\bar{\pi}_{123} \in \bar{Y}_1$, and $\bar{\pi}_{231'} \in \bar{Y}_2$.
12. Given a point $P \in \mathbb{R}^3$ and a plane $\pi \subset \mathbb{R}^3$, provide an algorithm for determining whether $P \in \pi$ or in which of the half spaces (“sides”) of π (partitions \mathbb{R}^3) P lies.
13. Generalize the 3-point-collinearity property for $N \geq 4$.

5.3 ♣ FT-3 Construction Algorithms

A dream mentioned at the outset is to do *multidimensional* synthetic constructions with this new coordinate system as has already been done for lines and now with planes and flats starting in \mathbb{R}^3 . Also, this is an additional opportunity to better understand the properties of the indexed points by indulging in the easy constructions they enable.

5.3.1 Planes and Lines

Half Spaces

With reference to Fig. 5.14, a plane ρ parallel to π intersects each of the two superplanes $\pi_1^s, \pi_{1'}^s$ at lines ℓ_ρ, ℓ'_ρ parallel to ℓ_π and $\ell_{\pi'}$ respectively. Hence the points $\bar{\pi}_{123}, \bar{\rho}_{123}$ are on a vertical line, and similarly, the points $\bar{\rho}_{231'}, \bar{\pi}_{231'}$ are also on a vertical line. Of course, this is entirely analogous to the conditions for parallel lines, as it should be, for after all, the triply indexed points represent lines (i.e., ℓ_π, ℓ_ρ , and ℓ'_π, ℓ'_ρ). From the representation of two parallel planes π, ρ , we agree that the \bar{H} ($H \in \mathbf{u}$) with the higher y -coordinate identifies the higher (above) plane. This clarifies how to distinguish and recognize half spaces as illustrated in Fig. 5.19, which is the cornerstone for the study of convexity. The two vertical lines together with \bar{H} also provide a coordinate system of planar coordinates as described in Section 5.1.1 based on the intersecting lines ℓ_π and $\ell_{\pi'}$. Usually there is no reason to distinguish between index permutations like 231' and 1'23, which are henceforth considered the same unless indicated otherwise.

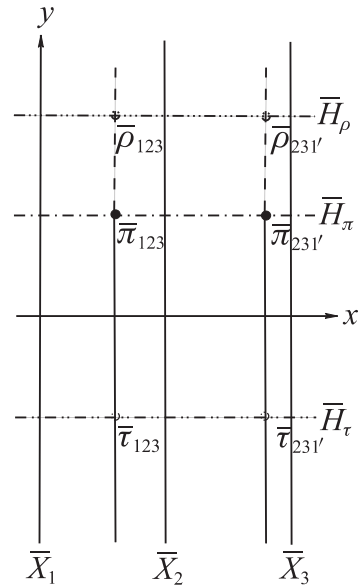


Figure 5.19. The parallel planes ρ, τ are above and below respectively the plane π . The upper half-space of π is marked by the two dashed half lines.

Line Contained in a Plane

Recognition that a line ℓ is contained in a plane π is immediate from the construction of the points $\bar{\pi}_{123}$. Specifically, a line $\ell \subset \pi$ if and only if $\bar{\ell}_{12}, \bar{\ell}_{23}, \bar{\pi}_{123}$ are

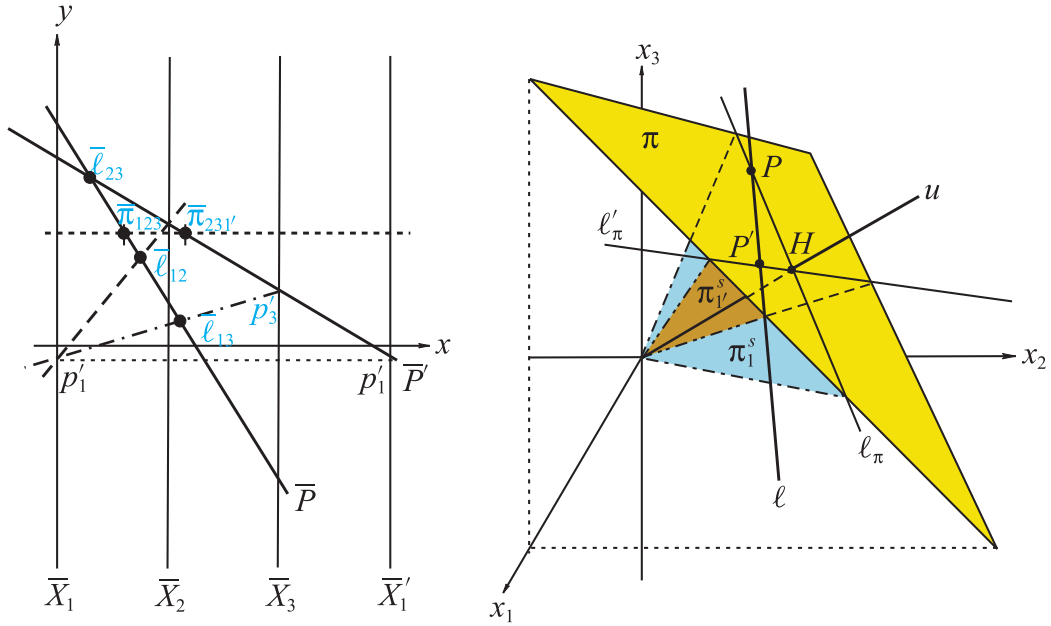


Figure 5.20. A line ℓ is contained in a plane $\pi \Leftrightarrow$ the points $\bar{\ell}_{12}, \bar{\ell}_{13}, \bar{\ell}_{23}, \bar{\pi}_{123}$ are on a line \bar{P} . Alternatively if $\bar{\ell}_{1'2}, \bar{\ell}_{1'3}, \bar{\ell}_{23}, \bar{\pi}_{231'}$ are on a line \bar{P}' , then $P = \ell_\pi \cap \pi$ and $P' = \ell_{\pi'} \cap \pi$.

collinear. This property is the equivalent of the collinearity in Fig. 5.5 for the two-vertical-lines planar representation. As illustrated in Fig. 5.20, the line $\ell \subset \pi$ intersects the superplanes at the two points $P = \ell \cap \ell_\pi$, $P' = \ell \cap \ell'_\pi$ with \bar{P}, \bar{P}' on $\bar{\pi}_{123}, \bar{\pi}_{231'}$ respectively, since these triply indexed points are actually $\bar{\ell}_\pi$ and $\bar{\ell}'_\pi$. Hence $\bar{\ell}_{23} = \bar{P} \cap \bar{P}'$, since the line \bar{P}' represents P' on the \bar{X}_2 -, \bar{X}_3 -, \bar{X}'_1 -axes. By transferring the value of its first coordinate p'_1 to the \bar{X}_1 -axis, we obtain the \bar{P}'_{12} portion of the *polygonal* line \bar{P}' in the \bar{X}_1 -, \bar{X}_2 -, \bar{X}_3 -axes with $\bar{\ell}_{12} = \bar{P} \cap \bar{P}'_{12}$. Therefore

$$\ell \subset \pi \Leftrightarrow \bar{\ell}_{12}, \bar{\ell}_{13}, \bar{\ell}_{23}, \bar{\pi}_{123} \in \bar{P}, \quad P = \ell_\pi \cap \pi. \quad (5.32)$$

The proof for the \Rightarrow direction is similar. Such a collinearity must also hold for the line \bar{P}' on $\bar{\ell}_{23}, \bar{\pi}_{1'23}$ for the \bar{X}_2 -, \bar{X}_3 -, \bar{X}'_1 -axes; that is, $\ell \subset \pi \Leftrightarrow \bar{\ell}_{1'2}, \bar{\ell}_{1'3}, \bar{\ell}_{23}, \bar{\pi}_{1'23}$ are on a line \bar{P}' where $P' = \ell'_{\pi'} \cap \pi$. This leads to a beautiful rotation \leftrightarrow translation duality, shown in Fig. 5.38 and presented shortly, analogous to that obtained with the vertical lines planar representation seen in Fig. 5.6.

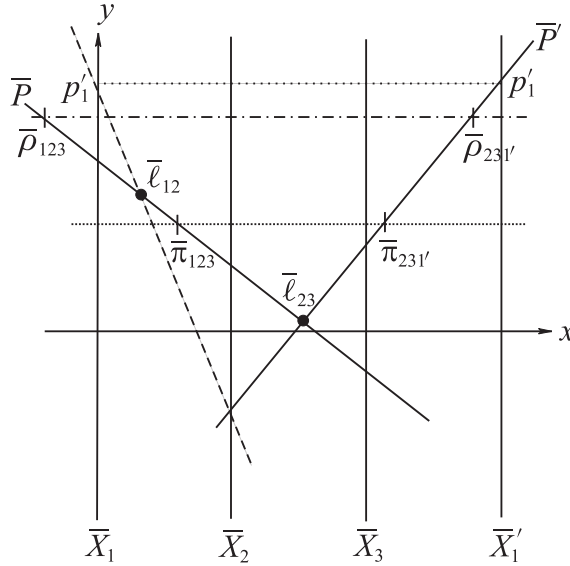


Figure 5.21. Two planes π and ρ intersecting at a line ℓ .

Intersecting Planes

To realize our goal of doing higher-dimensional constructions, that is, for $N > 3$, we gradually wean the reader from the need to use 3-D pictures. In the process we discover that many constructions even for 3-D are simpler to do and certainly to draw in \parallel -coords. Such is the construction shown in Fig. 5.21 for the intersection of two planes $\rho \cap \pi = \ell$. Finding the intersection $\pi \cap \rho \cap \pi_1^s = P$ is trivial, for \bar{P} is the line on the points $\bar{\pi}_{123}, \bar{\rho}_{123}$. Similarly for $\pi \cap \rho \cap \pi_{1'}^s = P'$, \bar{P}' is the line on the points $\bar{\pi}_{1'23}, \bar{\rho}_{1'23}$. Therefore $P, P' \in \ell = \pi \cap \rho$ and $\bar{\ell}_{23} = \bar{P} \cap \bar{P}'$. Transferring the p'_1 coordinate of P' to the \bar{X}_1 -axis provides the 12 portion of P' and its intersection with \bar{P} is the point $\bar{\ell}_{12}$. The points $\bar{\ell}_{12}, \bar{\ell}_{23}$ are both on $\bar{\pi}_{123}$ and $\bar{\rho}_{123}$, confirming that (5.32) the line ℓ is on both planes, simplicity itself!

5.3.2 The Four Indexed Points

The laborious construction for the four points $\bar{\pi}_{123}, \bar{\pi}_{231'}, \bar{\pi}_{31'2'}, \bar{\pi}_{1'2'3'}$ seen in Figure 5.17 can be greatly simplified. A 2-flat is determined by *any* three points it contains, and this is the basis in the ensuing construction algorithm. On each of the points $\bar{\pi}_{123}, \bar{\pi}_{231'}$, any two lines \bar{P} and \bar{P}' respectively are chosen as shown in Fig. 5.22. We know that $P \in \pi_1^s \cap \pi, P' \in \pi_{1'}^s \cap \pi$, and $\bar{\ell}_{23} = \bar{P} \cap \bar{P}'$, where $\ell \subset \pi$ is the line on P and P' . The p'_1 value (i.e., $\bar{P}' \cap \bar{X}_{1'}$) is transferred

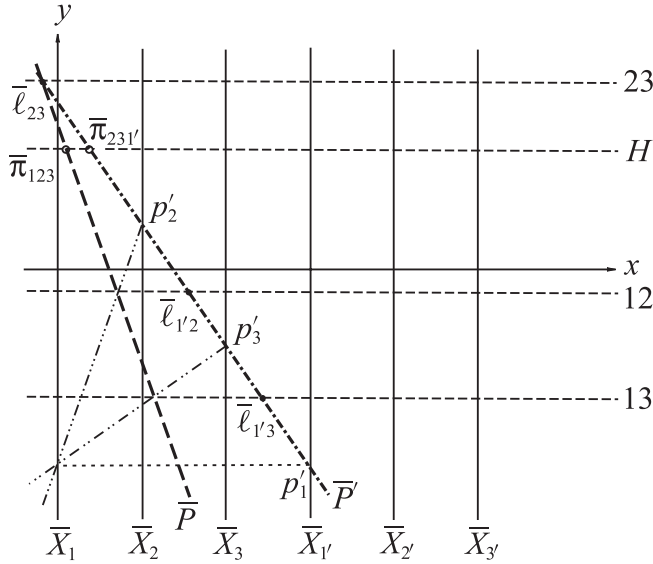


Figure 5.22. Representation of a plane $\pi \subset \mathbb{R}^3$ by two indexed points.

The first step in the construction of the points $\bar{\pi}_{32'1'}$, $\bar{\pi}_{1'2'3'}$ from $\bar{\pi}_{123}$, $\bar{\pi}_{231'}$. A (any) line $\ell \subset \pi$ is constructed as in Fig. 5.20. The points $\bar{\ell}_{12}$, $\bar{\ell}_{13}$ are constructed and the horizontal lines **12**, **13**, **23** are drawn on the $\bar{\ell}$'s with the corresponding indices.

from the $\bar{X}_{1'}$ to the \bar{X}_1 -axis and joined to the p'_2 coordinate intersecting \bar{P} at $\bar{\ell}_{12}$; the point $\bar{\ell}_{13}$ is similarly obtained. Horizontal lines **12**, **13**, **23** are drawn through the three $\bar{\ell}$ points as shown, providing at their intersections with the line \bar{P}' the points $\bar{\ell}_{1'2'}$, $\bar{\ell}_{1'3'}$ specifying ℓ on the \bar{X}_2 -, \bar{X}_3 -, $\bar{X}_{1'}$ -axes. Moving on to Fig. 5.23, the value p'_2 is transferred to the $\bar{X}_{2'}$ axis and joined to the p'_1 coordinate on $\bar{X}_{1'}$ intersecting the **12** line at the point $\bar{\ell}_{1'2'}$, which together with $\bar{\ell}_{1'3'}$ specifies ℓ on the \bar{X}_3 , $\bar{X}_{1'}$, $\bar{X}_{2'}$ axes. Therefore, as for P and P' , the line \bar{P}'' on these two points represents the point $P'' \in \pi_{1''}^S \cap \pi = \ell''_{\pi}$, where $\pi_{1''}^S$ is the superplane corresponding to this axes spacing (i.e., $d_{1'} = 3, d_{2'} = 4, d_3 = 2$). Hence $\bar{\pi}_{123} = \bar{P}'' \cap \bar{H}$ and $\bar{\ell}_{23} = \bar{P}'' \cap \mathbf{23}$. Proceeding as in Fig. 5.24, we place a line \bar{P}''' parallel to \bar{P} on the point $\bar{\ell}_{1'2'}$. In fact, P''' is the image of P in the $\bar{X}_{1'}$ -, $\bar{X}_{2'}$ -, $\bar{X}_{3'}$ axes, where, as we saw earlier, the superplane $\pi_{1'2'3'}^S$ is the image of π_{123}^S . The intersections of \bar{P}''' with \bar{H} , **13**, **23** complete the construction. The four points are referred to as the *augmented representation* of a plane. Use of the relations in (8.60) between the x -coordinates of the four indexed points simplifies the construction. By the way, Fig. 5.25, showing the $\bar{\ell}$ points corresponding to a line ℓ in the translated \parallel -coords systems, is obtained with a similar construction. The intersections of ℓ

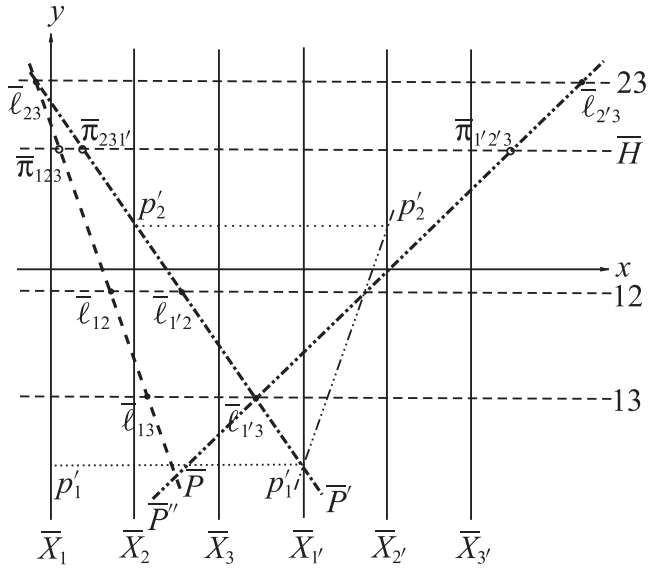


Figure 5.23. Construction of the third point $\bar{\pi}_{1'2'3}$.

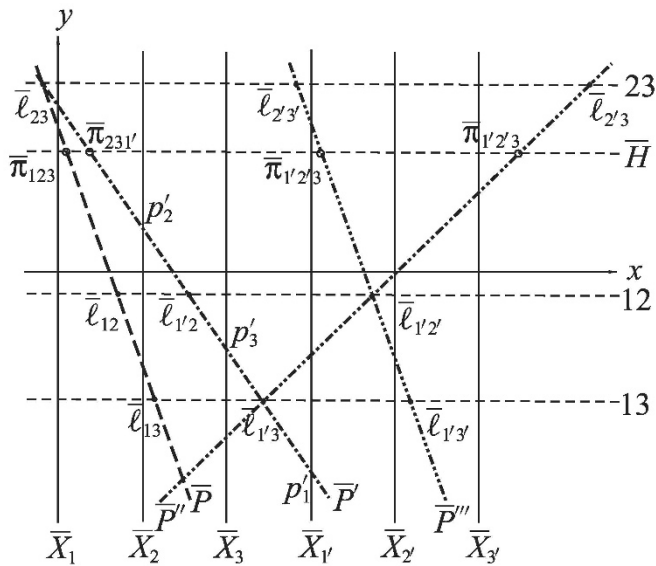


Figure 5.24. Construction of the fourth point $\bar{\pi}_{1'2'3'}$.

Note that \bar{P}''' on $\bar{\ell}_{1'2'}$ is parallel to \bar{P} , since P''' is the point $P = \pi \cap \ell_\pi$ on the $\bar{X}_{1'}$, $\bar{X}_{2'}$, $\bar{X}_{3'}$ axes.

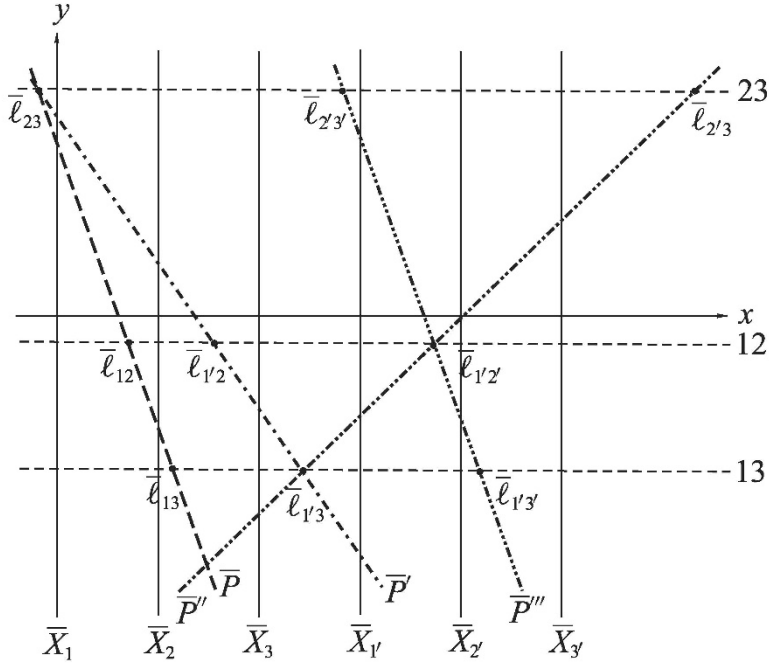


Figure 5.25. The collinear triples of points $\bar{\ell}$.

They represent a line ℓ in the four coordinate systems $\bar{X}_1 \bar{X}_2 \bar{X}_3$, $\bar{X}_2 \bar{X}_3 \bar{X}_{1'}$, $\bar{X}_3 \bar{X}_{1'} \bar{X}_{2'}$, and $\bar{X}_{1'} \bar{X}_{2'} \bar{X}_{3'}$.

with the superplanes are the points $P = \ell \cap \pi_1^s$, $P' = \ell \cap \pi_{1'}^s$, $P'' = \ell \cap \pi_{1'2'}^s$, $P''' = \ell \cap \pi_{1'2'3'}^s$. The further simplification of the construction via

$$x_{2'} = 6 - (x_{0'} + x_{1'}), \quad x_{3'} = 3 + x_{0'}, \quad (5.33)$$

obtained earlier, where $x_{2'}$, $x_{3'}$ are the x -coordinates of $\pi_{1'2'}^s$, $\pi_{1'2'3'}^s$, respectively, is left as an exercise.

♣ FT-3e

5.3.3 Special Planes

For future reference, we record the indexed points of some planes that are encountered on several occasions.

** The Principal 2-Planes

Our first examples are for the $x_1 x_2$ plane $\alpha : x_3 = 0$, $x_2 x_3$ plane $\beta : x_1 = 0$, and the $x_1 x_3$ plane $\gamma : x_2 = 0$. To obtain further practice in computing, the indexed points

of α are obtained from the fundamentals (i.e., as indexed points of lines) rather than from (5.16), (5.22), (5.24), (5.26). It is preferable to use the general relation

$$\ell : [a_1, a_2, a_3] \longrightarrow \bar{\ell} : (da_2, -a_3, a_1 + a_2),$$

rather than eq. (5.20) to obtain

$$\ell_\alpha = \alpha \cap \pi_1^s : \left\{ \begin{array}{l} \ell_{\alpha_{12}} : x_2 = \frac{1}{2}x_1 \\ \ell_{\alpha_{23}} : x_3 = 0 \end{array} \right\} \Rightarrow \left. \begin{array}{l} \bar{\ell}_{\alpha_{12}} = (\frac{1}{1-\frac{1}{2}}, 0, 1) \\ \bar{\ell}_{\alpha_{23}} = (1+1, 0, 1) \end{array} \right\} = (2, 0, 1) = \bar{\alpha}_{123}.$$

Proceeding as for (5.21), substituting the interaxis distance $d = -2$ between \bar{X}'_1 and \bar{X}_2 , and a translation by 3 from the y -axis yields

$$\begin{aligned} \ell'_\alpha = \alpha \cap \pi_{1'}^s : \left\{ \begin{array}{l} \ell'_{\alpha_{1'2}} : x_2 = -x_1 \\ \ell'_{\alpha_{23}} : x_3 = 0 \end{array} \right\} &\Rightarrow \left. \begin{array}{l} \bar{\ell}'_{\alpha_{1'2}} = (\frac{-2}{1+1} + 3, 0, 1) \\ \bar{\ell}'_{\alpha_{23}} = (1+1, 0, 1) \end{array} \right\} \\ &= (2, 0, 1) = \alpha_{231'}. \end{aligned}$$

For the third point, using the superplane given in (5.30) yields

$$\begin{aligned} \ell''_\alpha = \alpha \cap \pi_{1'2'}^s : \left\{ \begin{array}{l} \ell''_{\alpha_{1'2'}} : x_2 = 2x_1 \\ \ell''_{\alpha_{2'3}} : x_3 = 0 \end{array} \right\} &\Rightarrow \left. \begin{array}{l} \bar{\ell}''_{\alpha_{1'2'}} = (\frac{1}{1-2} + 3, 0, 1) \\ \bar{\ell}''_{\alpha_{2'3}} = (-2+4, 0, 1) \end{array} \right\} \\ &= (2, 0, 1) = \bar{\alpha}_{31'2'}. \end{aligned}$$

Hence the first three points are congruent, as they should be due to the zero coefficients of x_1 and x_2 in α 's equation. As has already been pointed out, the fourth superplane $\pi_{1'2'3'}^s$ coincides with the first π_1^s , though for the computation of the corresponding indexed point, the values based on the $\bar{X}'_1, \bar{X}'_2, \bar{X}'_3$ coordinate system are used, yielding

$$\begin{aligned} \ell'''_\alpha = \alpha \cap \pi_{1'2'3'}^s : \left\{ \begin{array}{l} \ell'''_{\alpha_{1'2'}} : x_2 = \frac{1}{2}x_1 \\ \ell'''_{\alpha_{2'3'}} : x_3 = 0 \end{array} \right\} &\Rightarrow \left. \begin{array}{l} \bar{\ell}'''_{\alpha_{1'2'}} = (\frac{1}{1-\frac{1}{2}} + 3, 0, 1) \\ \bar{\ell}'''_{\alpha_{2'3'}} = (1+4, 0, 1) \end{array} \right\} \\ &= (5, 0, 1) = \bar{\alpha}_{1'2'3'}. \end{aligned}$$

This point's distance of 3 from the others is, of course, due to the only nonzero coefficient in α 's equation. The representation is shown in Fig. 5.26 (left).

Alternatively, the location of the indexed points can be found graphically, by the intersection of two lines representing two points in the plane. The representations of β and γ are shown in Fig. 5.26 (right) and Fig. 5.27 respectively.

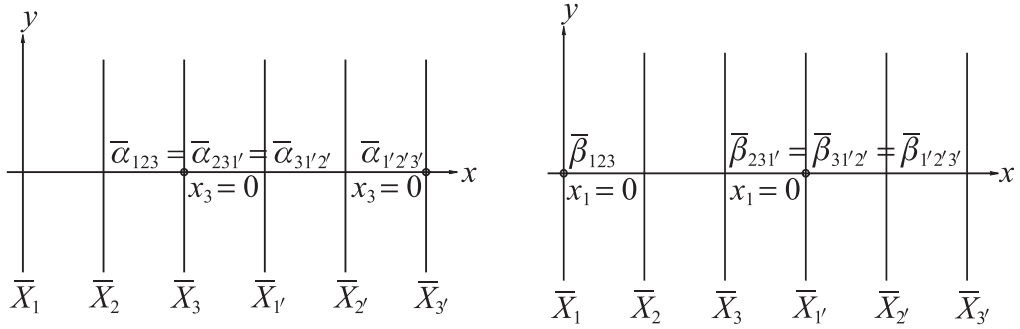


Figure 5.26. Indexed points corresponding to $\alpha : x_3 = 0$ for the x_1x_2 principal 2-plane on the left and for the x_2x_3 principal 2-plane $\beta : x_1 = 0$ on the right.

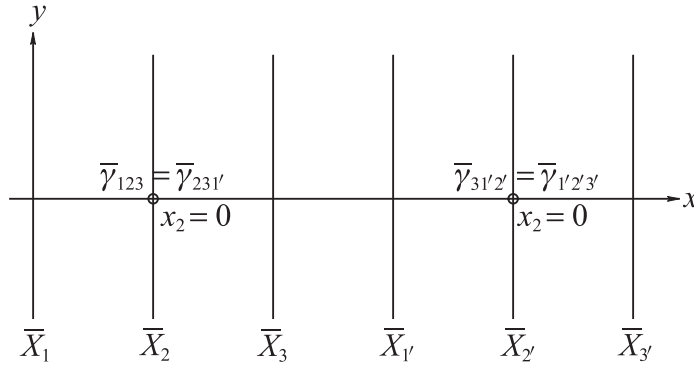


Figure 5.27. Indexed points corresponding to $\gamma : x_2 = 0$, the principal 2-plane x_1x_3 .

** The Constant Planes

Next are the more general planes $\kappa : x_i = k_0, i = 1, 2, 3$, with k_0 a constant. For

$$\kappa : x_1 = k_0 \quad (5.34)$$

via (5.16), (5.22), (5.24), (5.26), we obtain

$$\bar{\kappa}_{123} = (0, k_0, 1), \quad \bar{\kappa}_{231'} = \bar{\kappa}_{31'2'} = \bar{\kappa}_{1'2'3'} = (3, k_0, 1), \quad (5.35)$$

providing the representation shown in Fig. 5.28 (left). For

$$\kappa : x_2 = k_0, \quad (5.36)$$

the indexed points are

$$\bar{\kappa}_{123} = \bar{\kappa}_{231'} = (1, k_0, 1), \quad \bar{\kappa}_{31'2'} = \bar{\kappa}_{1'2'3'} = (4, k_0, 1), \quad (5.37)$$

for the representation in Fig. 5.28 (right). Continuing,

$$\kappa : x_2 = k_0 \quad (5.38)$$

has the indexed points

$$\bar{\kappa}_{123} = \bar{\kappa}_{231'} = \bar{\kappa}_{31'2'} = (2, k_0, 1), \quad \bar{\kappa}_{1'2'3'} = (5, k_0, 1), \quad (5.39)$$

with the representation in Fig. 5.29. Observe that the distance between adjacent points is three times the value of the corresponding coefficient.

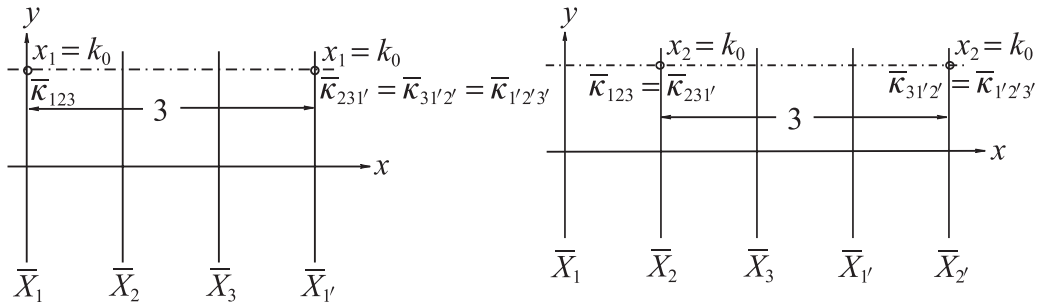


Figure 5.28. Indexed points representing constant planes.

For $\kappa : x_1 = k_0$ (left) and $\kappa : x_2 = k_0$ (right).

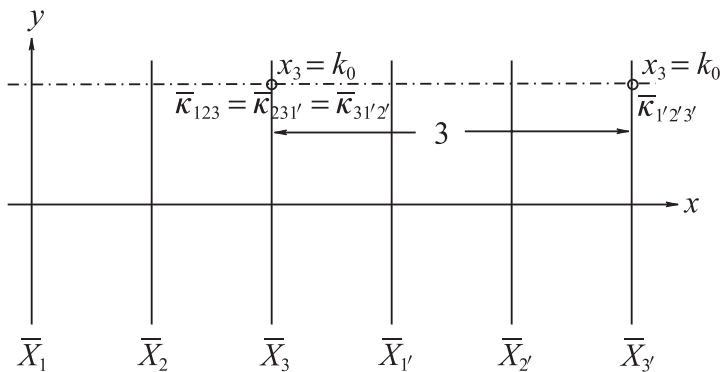


Figure 5.29. Representation of the plane $\kappa : x_3 = k_0$.

♣ FT-4 **Projecting Planes and Lines**

An important class of planes is the *projecting planes*, which are perpendicular to one of the principal 2-planes. Our first encounter with them was in Chapter 4, where in \mathbb{R}^N a line is defined as the intersection of $N - 1$ projecting planes. In general, the equation ℓ_{ij} for the pairwise linear relation between x_i and x_j describes a projecting plane perpendicular to the principal 2-plane $x_i x_j$. In Fig. 5.30 we see two projecting planes in \mathbb{R}^3 intersecting at a line ℓ . Where it is important to avoid an inconsistency, the projecting planes on α, β , and γ of a line ℓ are denoted by $\ell\alpha, \ell\beta$ instead of ℓ_{12}, ℓ_{23} respectively. Figure 5.31 is in fact Fig. 5.25 reincarnated in terms of the projecting planes of ℓ whose representation is directly discerned from the $\bar{\ell}$'s. For example, $\bar{\ell}\alpha_{31'2'} = \bar{\ell}\alpha_{1'2'3'}$, and these also coincide with $\bar{\ell}_{1'2'}$, since $\ell\alpha$'s x_3 coefficient is zero.

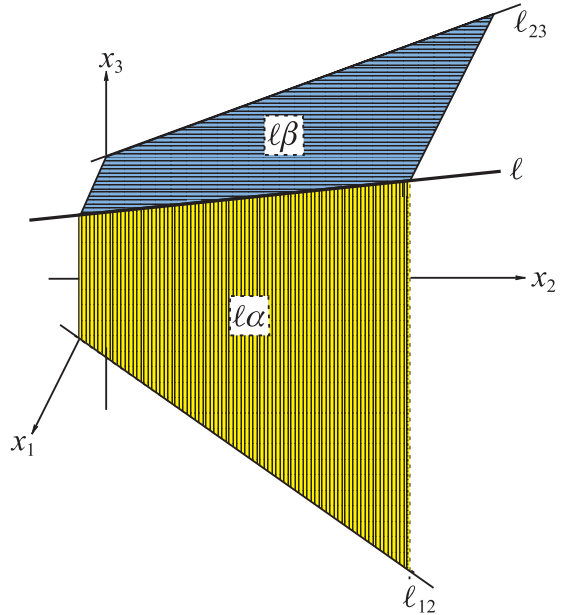


Figure 5.30. A line ℓ as the intersection of the projecting planes $\ell\alpha \perp \alpha$ and $\ell\beta \perp \beta$.

5.3.4 Intersecting a Plane with a Line

The description of a line ℓ in terms of projecting planes motivates an easy and intuitive algorithm, shown in Fig. 5.32, for finding its intersection with a plane π . The line $r = \ell\alpha \cap \pi$ is found, and then $R = r \cap \ell = \pi \cap \ell$. This is how it works. The algorithm's input is the initial data, consisting of the point pairs $\bar{\pi}_{123}, \bar{\pi}_{231'}$, and $\bar{\ell}_{12}, \bar{\ell}_{23}$, specifying the plane π and line ℓ respectively. The point $\bar{\ell}_{1'2'}$ is easily constructed as in Fig. 5.22 from the coordinates x_1, x_2 of *any* point on ℓ and transferring the x_1 value to the \bar{X}'_1 axis. The formality of writing $\bar{\ell}\alpha_{123} = \bar{\ell}_{12}$ and $\bar{\ell}\alpha_{231'} = \bar{\ell}_{1'2'}$ clarifies the use of the *planes* intersection construction in Section 5.3.1 to obtain $r = \pi \cap \ell\alpha$, actually \bar{r}_{23} , as in Fig. 5.3.3 with

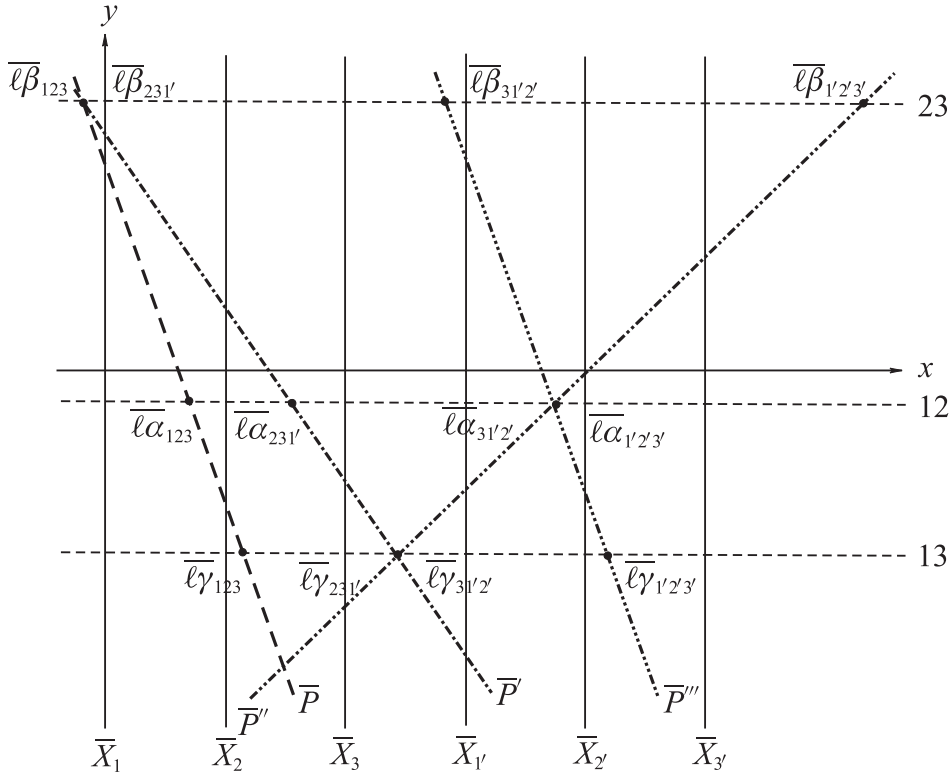


Figure 5.31. The projecting planes $\ell\alpha$, $\ell\beta$, $\ell\gamma$.
This is the line ℓ whose $\bar{\ell}$ points are shown in Fig. 5.25.

$\bar{r}_{12} = \bar{\ell}_{12}$, since $r \subset \ell\alpha$. Then $R = \ell \cap r = \ell \cap \pi$. In effect, the system of linear equations

$$\begin{cases} \pi & : c_1x_1 + c_2x_2 + c_3x_3 = c_0 \\ \ell_{1,2} & : c_{11}x_1 + c_{12}x_2 = c_{10} \\ \ell_{2,3} & : c_{22}x_2 + c_{23}x_3 = c_{20} \end{cases} \quad (5.40)$$

is solved geometrically. The construction is simple, and is easier to draw than in Cartesian coordinates. Also, it generalizes nicely to N dimensions.

♣ **FT-4e**

5.3.5 Points and Planes: Separation in \mathbb{R}^3

Let us return to the matter of half spaces, Section 5.3.1, now considering a plane π together with its normal vector \vec{n} partitioning \mathbb{R}^3 into two *oriented* half-spaces.

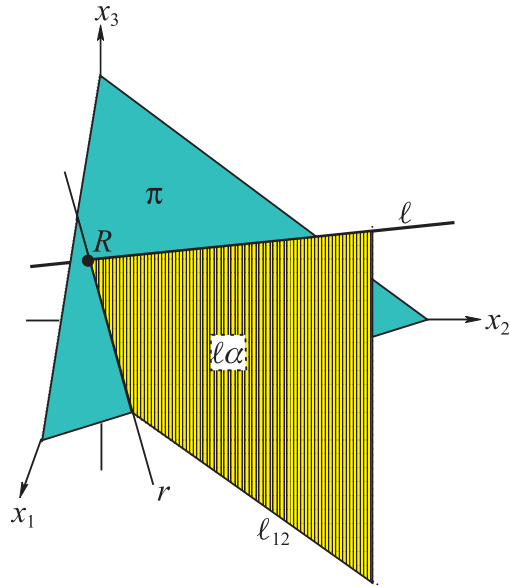


Figure 5.32. Finding $\ell \cap \pi$ using the projecting plane $\ell\alpha$.

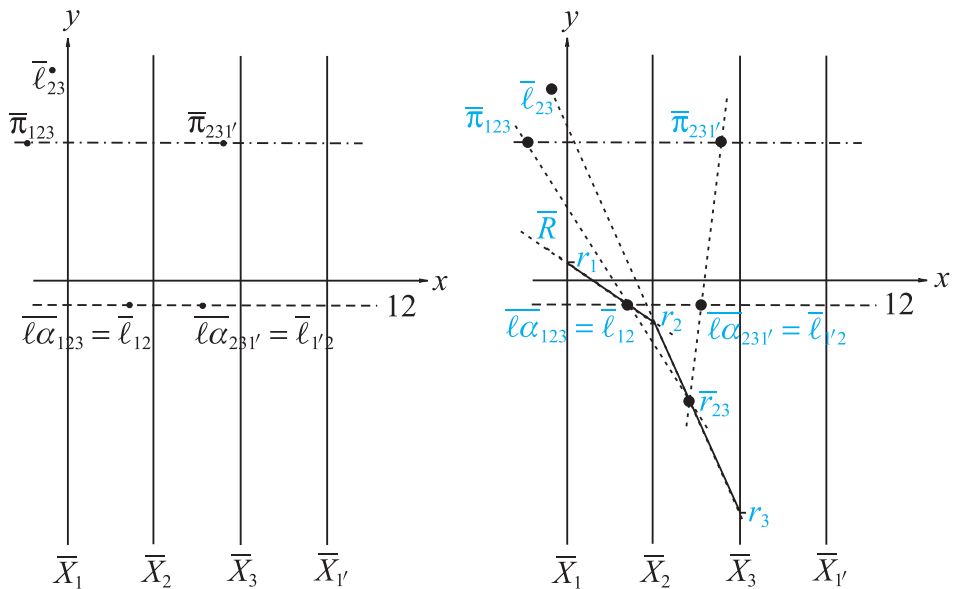


Figure 5.33. Intersection of a line with a plane.

On the left are the initial data, namely points specifying the plane π and line ℓ , for the intersection construction on the right. First $r = \pi \cap \ell\alpha$ is constructed (we need only \bar{r}_{23} , since $\bar{r}_{12} = \bar{\ell}_{12}$). Then $R = (r_1, r_2, r_3) = r \cap \ell = \pi \cap \ell$.

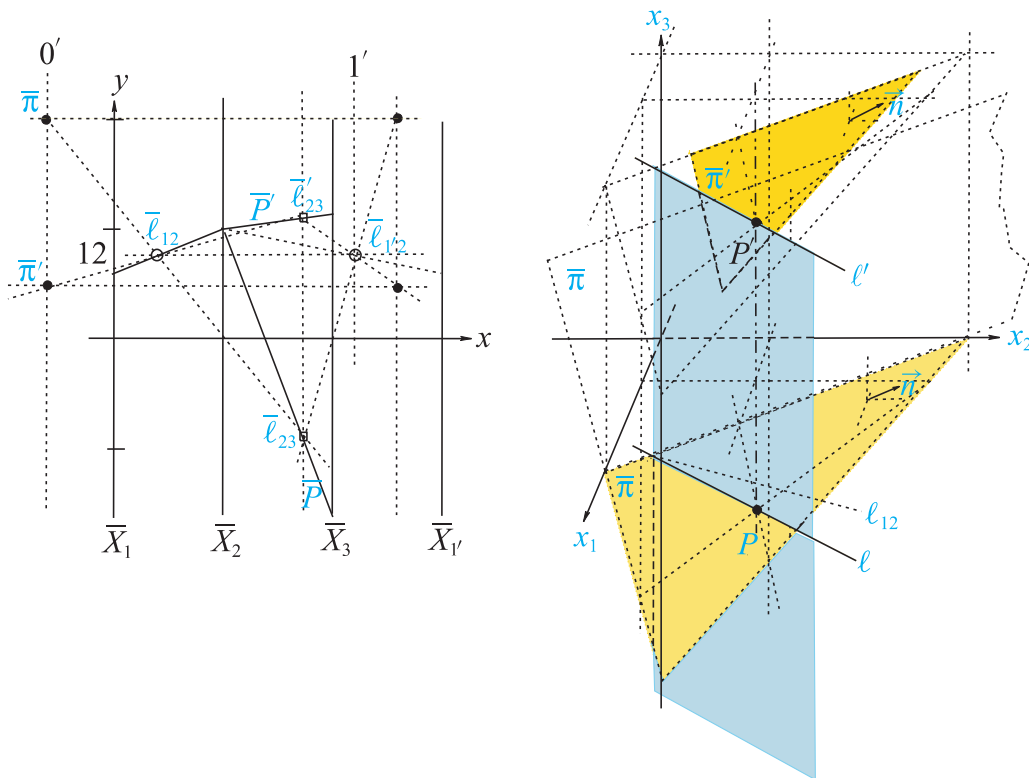


Figure 5.34. A point P on a plane π . The plane π' is constructed parallel to π and on a point P' which is *below* π .

Recall that a point P is “above” the plane π if it is on a plane π_P parallel to and above π as shown in Fig. 5.19. The normal vector \vec{n} points in the “above” direction.

The construction above instigates a “naïve” algorithm, benefiting from the special properties of $\|\text{-coords}$, for determining the relation of a point and a plane (i.e., on, above or below) rather than laboriously imitating the situation in Cartesian coordinates (as in the next section). The idea is simple and direct: given a plane π and a point P , take any line ℓ on P and intersect it with π . Then $P \in \pi \Leftrightarrow P = \ell \cap \pi$. This is shown in Fig. 5.34. A plane perpendicular to the x_1x_2 plane is chosen specifying the intersection of line ℓ_{12} with π , ℓ_{23} . The construction reveals that $P = \ell \cap \pi$. Taking another point $P' \notin \pi$ with the same x_1x_2 coordinates as P , the plane π' parallel to π and on P' is easily constructed. The point $\bar{\ell}'_{23}$ is placed at the intersection of \bar{P}' and, together with $\bar{\ell}_{12}$, specifies a line ℓ' parallel to ℓ .

Joining $\bar{\ell}_{12}$ to $\bar{\ell}'_{23}$ and extending until it intersects the $0'$ vertical line locates the point $\bar{\pi}'_0$, and similarly the point $\bar{\pi}'_1$ is found specifying π' —really simple! From the \parallel -coords display it is immediate that P' is *below*. One really has to struggle in order to understand that from the Cartesian coordinates, and a lot of care is needed to show this in that picture.

In *Linear Programming* the object is to move a plane parallel to another plane π (objective function) on the highest vertex of a convex polytope CP (constraints). It is worth trying to adapt this construction for many points to discover this top vertex.

5.3.6 ** Separation in \mathbb{R}^3 : An Old Approach

For²⁶ many applications it is useful to introduce *orientation* and discuss oriented half spaces, as shown in Fig. 8.11 (left) with respect to lines. A point $P : (p_1, p_2)$ is on, above, or below the line $\ell : x_2 = mx_1 + b$ if the expression $(p_2 - p_1m)$ is equal to, greater than or less than b , as shown on the right-hand part of Fig. 8.11. Since we are working in the projective plane \mathbb{P}^2 , we consider any regular (i.e., Euclidean) point as being below the ideal line ℓ_∞ . This situation is more “fun” in \parallel -coords, for there the separation criterion “flips” at $m = 1$. Correspondingly, \bar{P} is the line $y = (p_2 - p_1)x + p_1$, where to simplify matters we set $d = 1$. This is due to the *nonorientability* of \mathbb{P}^2 as already mentioned in Chapter 2 on geometry. For horizontal lines, “+” is the positive y direction, and for vertical lines, the positive x direction. The various cases for $\ell \neq \ell_\infty$ are summarized in the following lemma:

Lemma 5.3.1. P is on, below, above a line ℓ whose slope $m < 1 (m \geq 1) \iff \bar{P}$ is on, below (above), above (below) $\bar{\ell}$.

In \mathbb{R}^3 , for a plane π and a point P , for any plane ρ not parallel to π with $P \in \rho$, the intersection $\ell = \pi \cap \rho$ is found. *Within* the plane ρ , Lemma 5.3.1 is applied to determine whether P is on, below, or above $\pi \iff$ it is on, below, or above the line ℓ . The construction for this determination is shown in Fig. 5.36, where a point $T = (t_1, t_2, t_3)$ and plane π are shown. The constant plane $\kappa : x_1 = t_1$, which contains T and $\ell = \kappa \cap \pi$, is chosen as well as the point $P = (t_1, t_2, p_3) \in \ell$ (and hence in π). Viewing the picture in \parallel -coords, ℓ is found as the intersection of the two planes κ, π (by the construction in Fig. 5.21). Actually, only $\bar{\ell}_{23}$ is needed, since $\bar{\ell}_{12}$ coincides with $\bar{\kappa}_{123}$, so that κ is ℓ ’s projecting plane. The \bar{T}_{12} and \bar{P}_{12} portions of \bar{T}, \bar{P} coincide, and $\bar{\ell}_{23}$ is below \bar{T}_{23} and between the \bar{X}_2 and \bar{X}_3 axes, so the slope of ℓ_{23} is negative, so by Lemma 5.3.1, T is *above* π .

²⁶This section is only left for comparing the new and old approaches for solving this problem

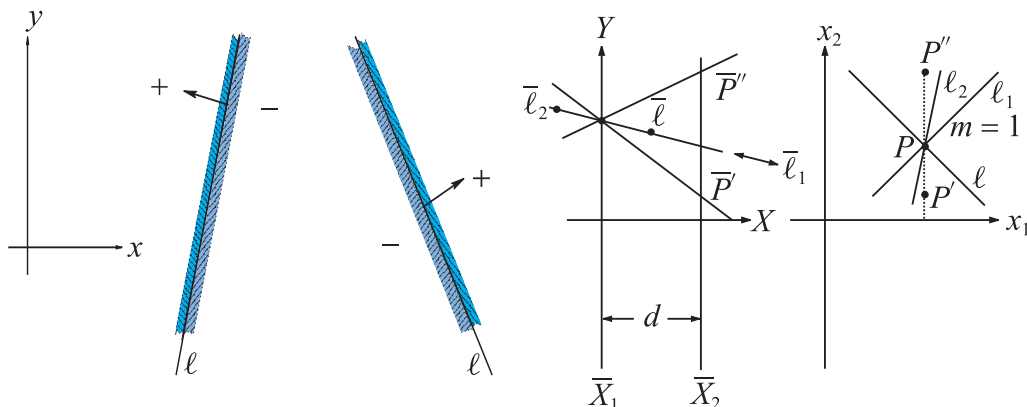


Figure 5.35. Oriented half spaces and orientation in \parallel -coords.

Left: Orienting half spaces on each side of a line $\ell : ax+by = c$ for $c \geq 0$. Points in “-” : $ax+by < c$ are “below” ℓ and “above” with the reverse inequality. Right: In \parallel -coords the above–below relation “flips” for lines with $m = 1$ due to the nonorientability of \mathbb{P}^2 .

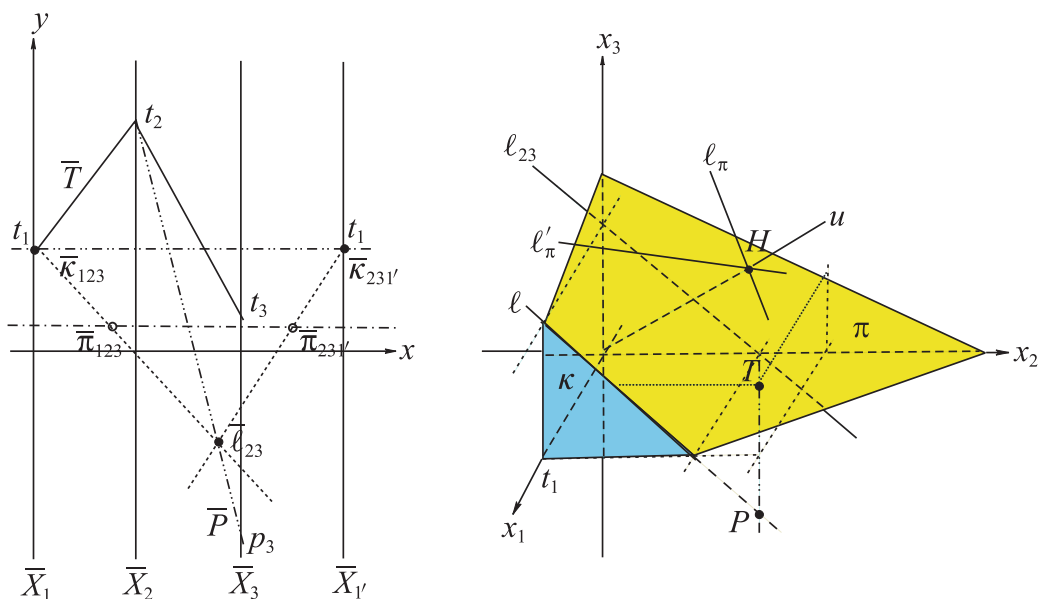


Figure 5.36. Point $T = (t_1, t_2, t_3)$ is above the plane π .

Line $\ell = \pi \cap \kappa$ where $\kappa : x_1 = t_1$ with $P = (t_1, t_2, p_3) \in \ell \cap \pi$. With $\bar{\ell}_{23}$ between the \bar{X}_2, \bar{X}_1 , i.e., the slope of $\bar{\ell}_{23}$ is negative, and below the portion \bar{T}_{23} of the polygonal line \bar{T} , T is above ℓ in the plane κ and also π . This is also clear from the picture since $T, P \in \kappa$ have the same x_1, x_2 coords and $p_3 < t_3$.

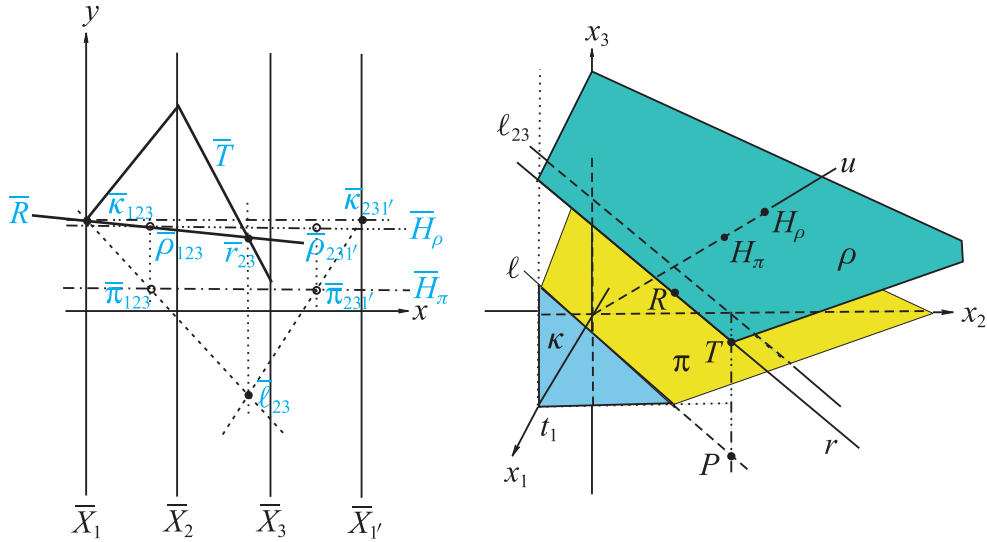


Figure 5.37. Point $T = (t_1, t_2, t_3)$ is above the plane π .
Line $\ell = \pi \cap \kappa$, where $\kappa : x_1 = t_1$ with $P = (t_1, t_2, p_3) \in \ell \cap \pi$.

While we are at it, we might as well find the plane ρ parallel to π and containing the point T . Through T the line r parallel to ℓ is determined (see Fig. 5.37) by the point \bar{r}_{23} at the intersection of the vertical line on $\bar{\ell}_{23}$ and \bar{T}_{23} . The line \bar{R} on \bar{r}_{23} and $\bar{r}_{12} = \bar{\ell}_{12}$ represents the point $R = r \cap \pi_1^s$ by Theorem 5.2.1 and is on the sought-after plane ρ , since $r \subset \rho$. The point $\bar{\rho}_{123}$ must be on \bar{R} at the intersection of the vertical line through $\bar{\pi}_{123}$, since ρ is to be parallel to π . All the rest is now determined; \bar{H}_ρ is the horizontal line through $\bar{\rho}_{123}$. It intersects the vertical line through $\bar{\pi}_{231'}$ at $\bar{\rho}_{231'}$.

With $\|\text{-}$ coords, then, we have the means to do easy synthetic constructions, such as those we enjoyed doing in elementary geometry, the lower-dimensional ones honing our intuition for their multidimensional counterparts, which we shortly meet.

5.3.7 ♣ FT-5 Rotation of a Plane about a Line and the Dual Translation

The analogue to the translation \leftrightarrow rotation duality in Fig. 5.6, based on the index-point representation, brings into play many of the constructions we just learned. Starting with *augmented representation* of a 2-flat with four rather than two points, as derived in Figs. 5.22, 5.23 and 5.24, the intent is to show the rotation of a plane $\pi^2 : c_1x_1 + c_2x_2 + c_3x_3 = c_0$ about a line π^1 . With only one line and plane

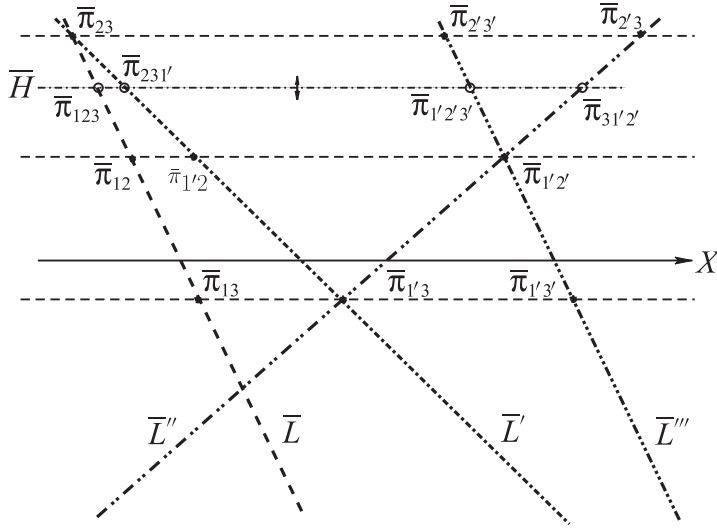


Figure 5.38. Rotation of a 2-flat (plane) about a 1-flat (line) in \mathbb{R}^3 .

This corresponds to a translation of the points with three indices on the horizontal line \bar{H} along the lines $\bar{L}, \bar{L}', \bar{L}'', \bar{L}'''$ joining the points with two indices.

appearing, the superscripts denoting the flats' dimensionality are not needed, for the flat's dimensionality can be determined from the number of subscripts. With reference to Fig. 5.38, a position of π^2 is shown by the points $\bar{\pi}_{123}, \bar{\pi}_{231'}, \bar{\pi}_{31'2'}, \bar{\pi}_{1'2'3'}$, representing π^2 on the horizontal line \bar{H} , while the line's π^1 representing points are $\bar{\pi}_{12}, \bar{\pi}_{13}, \bar{\pi}_{23}$ on the line \bar{L} , which, as in Fig. 5.25, also appears in terms of the other triples of axes with $L = \pi^1 \cap \pi_1^s$. Similarly, L', L'', L''' are the line's intersections with the second, third and fourth superplanes. To clarify,

$$\bar{\pi}_{1'2}\bar{\pi}_{1'3}\bar{\pi}_{2'3} \text{ on line } \bar{L}', \quad \bar{\pi}_{12'}\bar{\pi}_{13'}\bar{\pi}_{2'3'} \text{ on line } \bar{L}'', \quad \bar{\pi}_{12'}\bar{\pi}_{13'}\bar{\pi}_{2'3'} \text{ on line } \bar{L}'''.$$

With $\pi^1 \subset \pi^2$, the picture is that shown in Fig. 5.20, where the point $\bar{\pi}_{123}$ is on the intersection of \bar{H} with the line \bar{L} , and similarly,

$$\bar{\pi}_{231'} = \bar{H} \cap \bar{L}', \quad \bar{\pi}_{31'2'} = \bar{H} \cap \bar{L}'', \quad \bar{\pi}_{1'2'3'} = \bar{H} \cap \bar{L}'''. \quad (5.41)$$

The distance between adjacent points being proportional to the coefficient of the plane's equation (see Fig. 5.18), the intersections of the \bar{L} -lines mark the positions where coefficients are zero. There, the plane π^2 is perpendicular to the corresponding principal 2-plane. Specifically, at

$$\bar{H} \cap \bar{L} \cap \bar{L}', c_1 = 0 \text{ and } \pi^2 \perp x_2x_3 \text{ plane,}$$

$$\bar{H} \cap \bar{L}' \cap \bar{L}'', c_2 = 0 \text{ and } \pi^2 \perp x_1x_3 \text{ plane,} \quad (5.42)$$

$$\bar{H} \cap \bar{L}'' \cap \bar{L}''', c_3 = 0 \text{ and } \pi^2 \perp x_1x_2 \text{ plane.}$$

Now translate \bar{H} vertically with the four triply indexed $\bar{\pi}$ points moving along the four lines $\bar{L}, \dots, \bar{L}'''$. The conditions (5.41) hold, so that at *any* such translated position of \bar{H} , the containment property of Fig. 5.20 holds. Hence, the corresponding transformation must be the rotation of π^2 about π^1 passing along the way through the positions specified by (5.42) with all points on π^1 being invariant under the transformation. The variation of the coefficients can be followed throughout the transit.

An important phenomenon occurs when the horizontal line \bar{H} crosses the intersection points of the \bar{L}^i lines with $i = 0, 1, 2$ denoting the number of primes ('', ''', '''). The coefficients c_i are nonzero and, in this case, all positive when \bar{H} is in the position shown in Fig. 5.39. When \bar{H} is on the point $\bar{L}' \cap \bar{L}''$, the $\bar{\pi}_{231'}$, $\bar{\pi}_{31'2'}$ coincide. Hence $c_2 = 0$ (see Fig. 5.18) and the plane π is perpendicular to the x_1x_3 plane. Pursuing this closely, in Fig. 5.40, \bar{H} is just above this point, and in Fig. 5.41, just below the point, significantly showing that π has *flippe*. Consider the *oriented* plane together with its normal vector \vec{N} , whose components are the coefficients c_i . We show next that \bar{H} just traversing *all* three points $\bar{L}^i \cap \bar{L}^{i+1}$ corresponds to the plane making a 180° rotation about its axis, resulting in the normal $\vec{N} \rightarrow -\vec{N}$.

Fun and Games with the ILM

Open ILM2 and *click the **fift** button with the ||-coords icon.*
The plane and line in Cartesian coordinates are on the right.
A picture similar to Fig. 5.38 appears on the left.

For some experimentation:
 translate vertically the horizontal line \bar{H} on the four $\bar{\pi}$ points and observe
 that the $\bar{\pi}$ slide along the $\bar{L}, \dots, \bar{L}'''$ lines.
 Note the distance between the adjacent $\bar{\pi}$
 and the coefficients of the plane's equation at the bottom.

In preparation for the next discussion,
 place the line \bar{H} on each of the line intersections, i.e., $\bar{L} \cap \bar{L}'$, etc.
 Note that then the plane is perpendicular to a principal 2-plane.

Change the line and/or plane and repeat.

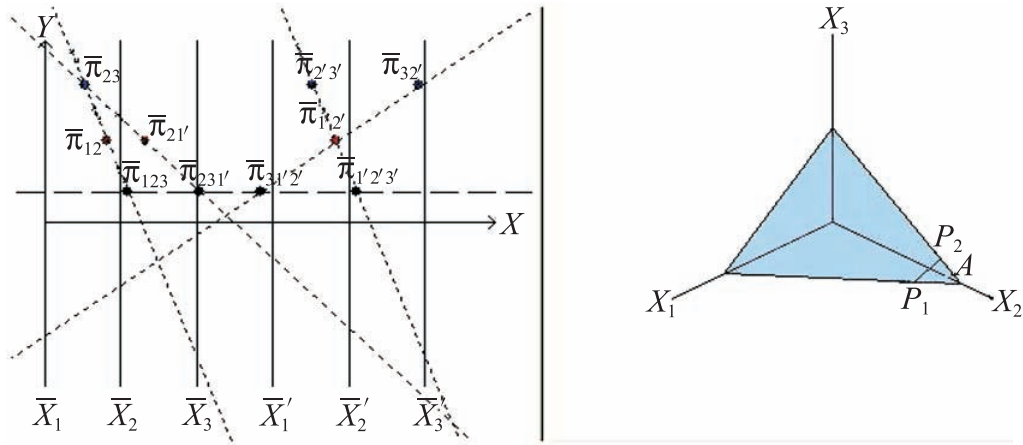


Figure 5.39. Rotation of a plane about a line in Cartesian and \parallel -coords. In this instance, all the coefficients c_i of the plane's equation are less than or equal to zero.

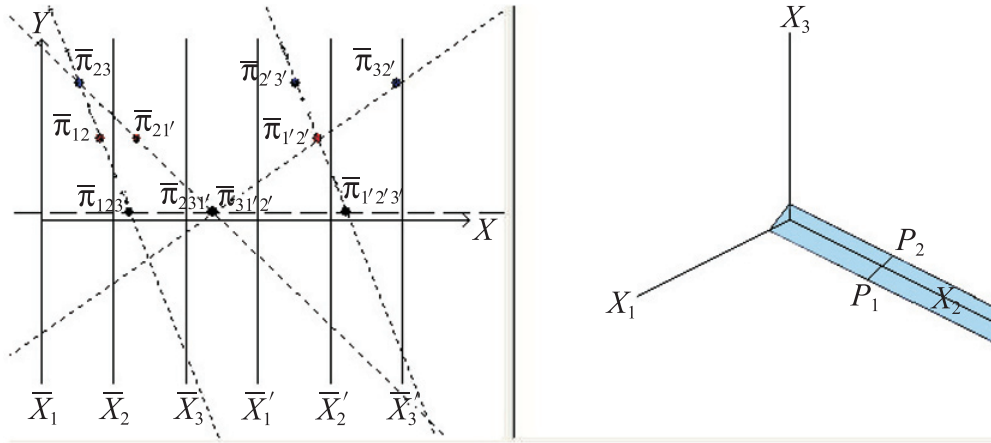


Figure 5.40. Here c_2 is small and positive. The line \bar{H} being just above the point $\bar{L}' \cap \bar{L}''$ indicates that the plane is nearly perpendicular to the 2-plane x_1x_3 .

Let

$$\bar{L}^i : y = m_i x + b_i, i = 0, 1, 2, 3, \quad (5.43)$$

$$(x_I, y_I) = \bar{L}' \cap \bar{L}'' = \left(-\frac{(b_2 - b_1)}{(m_2 - m_1)}, \frac{(m_2 b_1 - m_1 b_2)}{(m_2 - m_1)} \right). \quad (5.44)$$

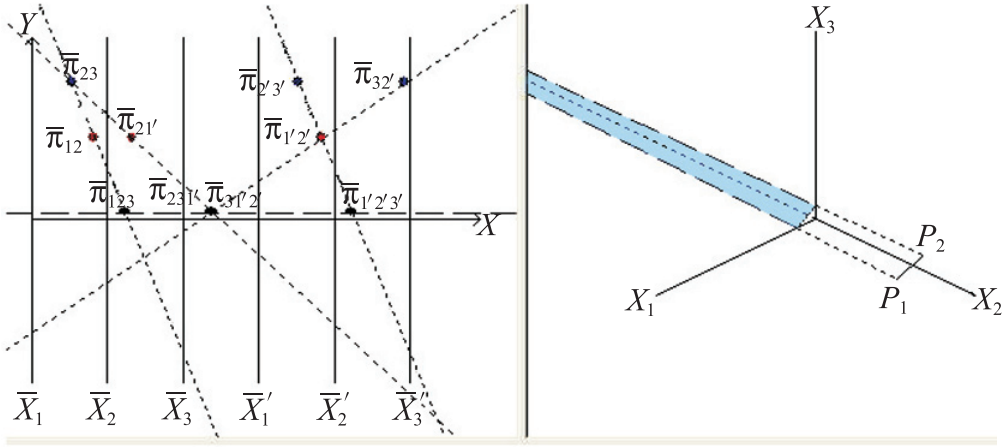


Figure 5.41. Here c_2 is small and negative.

The line \bar{H} is just below the point $\bar{L}' \cap \bar{L}''$. The plane is nearly perpendicular to the principal plane x_1x_3 but has “flipped” compared to the previous figure.

For a value $\epsilon > 0$, with the superscripts $+$, $-$ denoting values above or below y_I respectively,

$$y^+ = y_I + \epsilon = m_i x_i^+ + b_i, \quad y^- = y_I - \epsilon = m_i x_i^- + b_i,$$

$$x_i^+ = x_I + \epsilon/m_i, \quad x_i^- = x_I - \epsilon/m_i.$$

Then

$$c_2^+ = k(x_2^+ - x_1^+) = \epsilon(m_1 - m_2)/m_1 m_2,$$

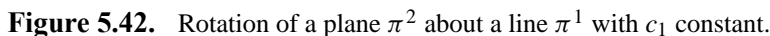
$$c_2^- = k(x_2^- - x_1^-) = -\epsilon(m_1 - m_2)/m_1 m_2,$$

$$\Rightarrow \quad c_2^+/c_2^- = -1, \quad (5.45)$$

where k is a proportionality constant. It is clear from Fig. 5.39 or from an easy calculation that $c_1^+/c_1^- \approx 1 \approx c_3^+/c_3^-$ for ϵ small. Therefore, as $\epsilon \rightarrow 0$,

$$c_1^+/c_1^- \rightarrow 1, \quad c_3^+/c_3^- \rightarrow 1, \quad \text{and} \quad c_2^+/c_2^- \rightarrow -1. \quad (5.46)$$

As \bar{H} traverses the point $\bar{L}' \cap \bar{L}''$, the plane π remains perpendicular to the x_1x_3 plane, with the x_2 component of the normal \bar{N} flipping its orientation by 180° . With c_1 and c_3 unchanged, this amounts to the plane π “flipping” with respect to the principal plane x_1x_3 . That is what is seen in Figs. 5.40, 5.41. Of course, a



Corollary 5.3.2 (Flip) *For an oriented plane π having normal vector $\vec{N} = (c_1, c_2, c_3)$ rotating about a line ℓ , the traversal of the point $\bar{L}^i \cap \bar{L}^{i+1}$, $i = 0, 1, 2$, by \bar{H} corresponds to $c_i^+/c_i^- \rightarrow -1$, $c_j^+/c_j^- \rightarrow 1$, $j \neq i$.*

Specializations and Generalizations

We have seen that the rotation of a plane about an axis is completely determined by the translation of \bar{H} with the $\bar{\pi}_{123}, \bar{\pi}_{231'}$ along the two lines \bar{L}, \bar{L}' . Pictorializing the rotation facilitates its customization to specific cases. For example, what rotation

leaves the coefficient c_1 of the plane's equation invariant? The geometric requirement is for \bar{L} and \bar{L}' to be parallel, that is, their point of intersection $\bar{\pi}_{23}$ is ideal. The rotation then leaves the first coefficient c_1 invariant. The direction of \bar{L} and hence of $\bar{\pi}_{23}$ is determined by b_3 . All of this is seen in Fig. 5.42 with the corresponding translation of \bar{H} . The $\bar{\pi}_{123}, \bar{\pi}_{231'}, \bar{\pi}_{31'2'}, \bar{\pi}_{1'2'3'}$ rolling along $\bar{L}, \bar{L}', \bar{L}'', \bar{L}'''$ respectively show the variation of the other coefficients of the plane's equation. The construction starts by choosing \bar{L} and \bar{L}' in the direction given by b_3 , while the horizontal distance is $3c_1$. As in Fig. 5.22, the coordinates of \bar{L}' determine the points $\bar{\pi}_{12}, \bar{\pi}_{13}$ and the lines **12**, **13**. Or conversely, these are obtained from the line's equation, either way proceeding as outlined in Section 5.3.2, where only \bar{L}'' is not parallel to \bar{L} . The pencil of planes obtained by the vertical translation of \bar{H} is now

$$\pi^2 : (x_3 - x_2 - b_3) + k(x_2 - m_2x_1 - b_2) = 0. \quad (5.47)$$

Here it is assumed that the axis of rotation π^1 lies on the plane π^2 (see Exercise 8). Then for $\pi^2 : c_1x_1 + c_2x_2 + c_3x_3 = c_0$,

$$\begin{aligned} \pi_{23}^1 : x_3 &= x_2 + b_3, \\ \pi_{12}^2 : x_2 &= \frac{c_1}{c_2+c_3}x_1 + \frac{c_0-c_3b_3}{c_2+c_3}. \end{aligned} \quad (5.48)$$

The translation of \bar{H} is not the essential “ingredient,” for there is no vertical movement when the axis of a plane's rotation is on the point H . For our purposes this means that the lines \bar{L}, \bar{L}' are on \bar{H} and the points $\bar{\pi}_{123}, \bar{\pi}_{231'}$ move only horizontally, and there are some interesting subcases. For example, rotation of a plane π about its ℓ_π results in $\bar{\pi}_{123}$ being stationary (since it is ℓ_π) while $\bar{\pi}_{231'}$ is translated along \bar{H} , together, of course, with the other two points. Similarly, rotation about ℓ'_π leaves $\bar{\pi}_{231'}$ stationary. Rotation about a general axis $\pi^1 \subset \pi^2$ on H results in all triply indexed points moving simultaneously along \bar{H} .

The rotation of a plane π viewed in \parallel -coords as the movement of points opens up new vistas for generalization. For parallel translations of π , the corresponding movement in \parallel -coords is vertical, with the “pencil” of planes so generated being the family represented in Fig. 5.19. It is useful to design the transformation of planes defining the corresponding pair of loci of the points $\bar{\pi}_{123}, \bar{\pi}_{231'}$ on the plane. For example, in Chapter 9 it is shown that any developable surface σ is represented by a pair of curves with the same minimum and maximum y -coordinate, σ being the envelope of a one-parameter family of tangent planes (Exercise 9). So in \mathbb{R}^N , a developable hypersurface is represented by $N - 1$ such curves.

With some care, the composition of transformations can be specified and visualized as the union of the corresponding loci. Consider, for example, the representation of “wobbling,” as in the motion of a spinning top, as the composition of a rotation about an axis rotating about its tip while moving along a path in a plane (Exercise 10). The interesting generalizations to \mathbb{R}^N are an excellent research topic; see, for example, [15, pp. 403–409], and Sections 2.7 and 2.8.

Exercises

1. Given three noncollinear points $P \in \pi_1^s$, $P' \in \pi_{1'}^s$, $P'' \in \pi_{1''}^s$, provide an algorithm that finds the representation of the 2-flat π^2 that contains the three points. Even though a plane is uniquely specified by three noncollinear points, in this case the algorithm should take advantage of the fact that three given points belong to the three principal sp .
2. For the construction of the four indexed points in Section 5.3.2, prove that any choice of the three lines \bar{L} , \bar{L}' , \bar{L}'' gives the same result.
3. Following the representation of a line as the intersection of two projection planes,
 - (a) express a point P as the intersection of three constant planes (as determined by its coordinates), and
 - (b) for any plane π and point P , determine if P is above or on the plane π .
4. Draw the figures corresponding to Fig. 5.42 but for rotations leaving the second and separately the third coefficients of the plane’s equation invariant.
5. Find k in (5.47) as a function of the corresponding y -coordinate of \bar{H} and then specialize it to the subcases given.
6. Draw the figure corresponding to Fig. 5.38 for an axis of rotation through the point H of the plane. In this case \bar{L} and \bar{H} coincide.
7. Find the angle of rotation of a 2-flat π about a 1-flat π^1 as a function of the motion of the triply indexed $\bar{\pi}$ points. Specialize to the case that π^1 is on the point H .
8. Draw the figure showing the rotation of a plane π^2 about a line π^1 *not* on π^2 and include $\pi^2 \cap \pi^1$.
9. Describe the transformation of a plane in which the points $\bar{\pi}_{123}$ and $\bar{\pi}_{1'23}$ each trace two different ellipses having the same minimum and maximum y -coordinate.
10. Provide the geometric locus for the representation of a plane’s wobbling (last sentence of this section).
11. Construct the constant-planes representation in higher dimensions.

12. Prove the correctness of the algorithm for

- (a) The intersection of two planes.
- (b) The intersection of a line with a plane.

This means that you should show that for any input, the algorithm terminates with the correct output.

5.4 Hyperplanes and p -Flats in \mathbb{R}^N

5.4.1 The Higher-Dimensional Superplanes

Just as for \mathbb{R}^3 , we expect the superplanes in \mathbb{R}^N consisting of points of the form (5.7) to play a fundamental role. They are denoted by π^{Ns} to mark their dimensionality N , abbreviated to sp , and are on the line

$$u : x_2 = x_1, x_3 = x_2, \dots, x_i = x_{i-1}, \dots, x_N = x_{N-1}. \quad (5.49)$$

The axes spacing $\mathbf{d} = (d_1, d_2, \dots, d_i, \dots, d_N)$ notation is used, designating the \bar{X}_i axes placed at $x_i = d_i$. Being 2-flats, the sp are described by $N - 2$ linearly independent equations, which after some matrix manipulations can be written with two variables each as

$$\pi^{Ns} : \begin{cases} \pi_{123}^{Ns} : (x_2 - x_1) + k_1(x_3 - x_2) = 0, \\ \pi_{234}^{Ns} : (x_3 - x_2) + k_2(x_4 - x_3) = 0, \\ \dots \\ \pi_{i(i+1)(i+2)}^{Ns} : (x_{i+1} - x_i) + k_i(x_{i+2} - x_{i+1}) = 0, \\ \dots \\ \pi_{(N-2)(N-1)N}^{Ns} : (x_{N-1} - x_{N-2}) + k_{N-1}(x_N - x_{N-1}) = 0. \end{cases} \quad (5.50)$$

The form of the equations shows that in the 3-dimensional subspace $x_{i-2}x_{i-1}x_i$, the $\pi_{(i-2)(i-1)i}^{Ns}$ are the pencil of 2-flats on the 1-flat (line) $x_{i-1} = x_{i-2}, x_i = x_{i-1}$. Noting that each such 2-flat contains the points $(1, 1, 1), (d_{i-2}, d_{i-1}, d_i)$ enables the elimination of the parameter k_i in the equations π^{Ns} . Rewriting in terms of the axes spacing yields

$$\begin{aligned}
\pi^{Ns} (d_1, d_2, d_3) : (d_3 - d_2)x_1 + (d_1 - d_3)x_2 + (d_2 - d_1)x_3 &= 0, \\
\pi^{Ns} (d_2, d_3, d_4) : (d_4 - d_3)x_2 + (d_2 - d_4)x_3 + (d_3 - d_2)x_4 &= 0, \\
\ldots \\
\pi^{Ns} (d_i, d_{i+1}, d_{i+2}) : (d_{i+2} - d_{i+1})x_i + (d_i - d_{i+2})x_{i+1} \\
&+ (d_{i+3} - d_i)x_{i+2} = 0, \\
\ldots \\
\pi^{Ns} (d_{N-2}, d_{N-1}, d_N) : (d_N - d_{N-1})x_{N-2} + (d_{N-2} - d_N)x_{N-1} \\
&+ (d_{N-1} - d_{N-2})x_N = 0.
\end{aligned} \tag{5.51}$$

For example, in \mathbb{R}^4 the sp are given by

$$\pi^{4s} : \begin{cases} \pi_{123}^{4s} : (d_3 - d_2)x_1 + (d_1 - d_3)x_2 + (d_2 - d_1)x_3 = 0, \\ \pi_{234}^{4s} : (d_4 - d_3)x_2 + (d_2 - d_4)x_3 + (d_3 - d_2)x_4 = 0. \end{cases} \tag{5.52}$$

This is a good time to define a recursive notational convention²⁷ for the axis spacing resulting from the successive translations of the \bar{X}_i to the new positions $\bar{X}_{i'}$. For the initial N -axes system $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N$, we write

$$\mathbf{d}_N^0 = \overbrace{(0, 1, 2, \dots, i-1, \dots, N-1)}^i,$$

stating that the first axis \bar{X}_1 is placed at $x = 0$, \bar{X}_2 is placed at $x = 1, \dots$ and \bar{X}_N at $x = N - 1$. After translating the \bar{X}_1 axis one unit to the right of \bar{X}_N and renaming it $\bar{X}_{1'}$, the axis spacing N -tuple is

$$\mathbf{d}_N^1 = \mathbf{d}_N^0 + (N, 0, \dots, 0) = \overbrace{(N, 1, 2, \dots, i-1, \dots, N-1)}^i.$$

And after i such successive unit translations with the \bar{X}_i axis in position $\bar{X}_{i'}$, the axis spacing for $i = 0, \dots, N, k = 1, \dots, N$ is given by

$$\begin{aligned}
\mathbf{d}_N^i &= \mathbf{d}_N^0 + \overbrace{(N, \dots, N, 0, \dots, 0)}^i \\
&= \overbrace{(N, N+1, \dots, i-1+N)}^i, i, \dots, N-1 = (d_{ik}).
\end{aligned} \tag{5.53}$$

²⁷I am indebted to Liat Cohen for proposing this notation.

To clarify the indices, i refers to the number of the axis translations, and k is the position (component) within the vector \mathbf{d}_N^i . By means of the step function

$$S_i(k) = \begin{cases} 1 & i \geq k, \\ 0 & i < k, \end{cases} \quad (5.54)$$

the axis spacing after the i th translation is

$$\mathbf{d}_N^i = (d_{ik}) = (k - 1 + N S_i(k)). \quad (5.55)$$

When the dimensionality is clear from the context, the subscript N can be omitted. For a flat π^P expressed in terms of the \mathbf{d}_N^i spacing, the points $\bar{\pi}_{1', \dots, i', i+1, \dots, N}^P$ of its representation are denoted compactly by $\bar{\pi}_{i'}^P$, and it is consistent to write $\pi^P = \pi_{0'}^P$, that is, π^P described in terms of the axis spacing \mathbf{d}_N^0 .

Returning now to the higher-dimensional sp , in \mathbb{R}^4 for the standard spacing \mathbf{d}^0 ,

$$\pi_{0'}^{4s} : \begin{cases} \pi_{123}^s : x_1 - 2x_2 + x_3 = 0, \\ \pi_{234}^s : x_2 - 2x_3 + x_4 = 0. \end{cases} \quad (5.56)$$

The axes are translated, as in \mathbb{R}^3 , to generate different sp corresponding to rotations about u , which are summarized in Fig. 5.43. First, the axis \bar{X}_1 is translated to position

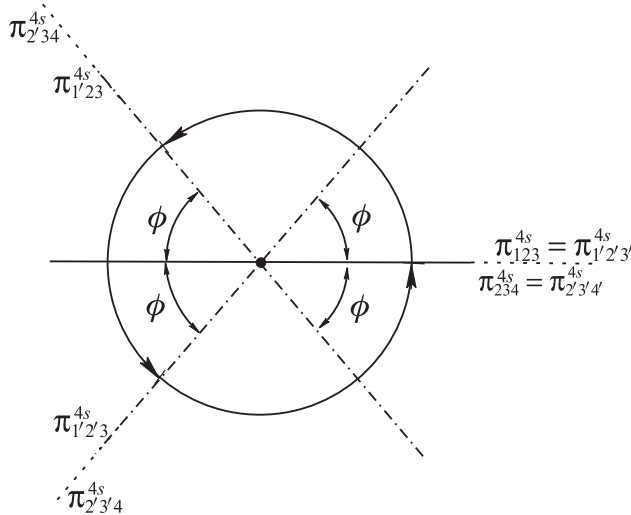


Figure 5.43. The rotations of π^{4s} about the line u .

This is a projection on a plane perpendicular to u , and so projections of the 3-flats $\pi_{123}^{4s}, \pi_{234}^{4s}$ are lines.

$\bar{X}_{1'}$ one unit to the right of \bar{X}_4 with the resulting axis spacing $\mathbf{d}^1 = (4, 1, 2, 3)$, yielding

$$\pi_{1'}^{4s} : \begin{cases} \pi_{1'23}^s : x_1 + 2x_2 - 3x_3 = 0, \\ \pi_{234}^s : x_2 - 2x_3 + x_4 = 0. \end{cases} \quad (5.57)$$

The angle of rotation between π_{123}^{4s} and $\pi_{1'23}^{4s}$ computed via (5.29) is $\cos^{-1}(-\sqrt{3/7}) = 180^\circ - \phi$, $\phi = \cos^{-1}(\sqrt{3/7}) \approx 49.1^\circ$. Note that since π_{234}^s remains unchanged, this is not a complete rotation of π_{1234}^{4s} about the line u . Proceeding with the translation of \bar{X}_2 to position \bar{X}_2' one unit to the right of $\bar{X}_{1'}$ provides the axis spacing $\mathbf{d}^2 = (4, 5, 2, 3)$ and the sp

$$\pi_{2'}^{4s} : \begin{cases} \pi_{1'2'3}^s : 3x_1 - 2x_2 - x_3 = 0, \\ \pi_{2'34}^s : x_2 + 2x_3 - 3x_4 = 0. \end{cases} \quad (5.58)$$

The angle between $\pi_{1'23}^s$ and $\pi_{1'2'3}^s$ is $\cos^{-1}(-1/7) = 2\phi$, while the angle between π_{234}^s and $\pi_{2'34}^s$ is $180^\circ - \phi$. With the translation of \bar{X}_3 to position $\bar{X}_{3'}$ one unit to the right of \bar{X}_2' , $\mathbf{d}^3 = (4, 5, 6, 3)$ and

$$\pi_{3'}^{4s} : \begin{cases} \pi_{1'2'3'}^s : x_1 - 2x_2 + x_3 = 0, \\ \pi_{2'3'4}^s : 3x_2 - 2x_3 - x_4 = 0, \end{cases} \quad (5.59)$$

returning $\pi_{1'2'3'}^s$ to its original position π_{123}^s while bringing $\pi_{2'3'4}^s$ to an angle 2ϕ from $\pi_{2'34}^s$. Again this is not a complete rotation of the whole sp about the line u . The final translation of \bar{X}_4 to $\bar{X}_{4'}$ one unit to the right of $\bar{X}_{3'}$ provides $\mathbf{d}^4 = (4, 5, 6, 7)$ and

$$\pi_{4'}^{4s} : \begin{cases} \pi_{1'2'3'}^s : x_1 - 2x_2 + x_3 = 0, \\ \pi_{2'3'4'}^s : x_2 - 2x_3 + x_4 = 0, \end{cases} \quad (5.60)$$

which is identical to π_{1234}^s . Unlike the case in \mathbb{R}^3 , here the rotations' angles are not all equal, though the sum is a full circle. The ‘‘anomaly’’ suggests that $\phi(N)$ is a function of the dimensionality N with $\phi(3) = 60^\circ$ and $\phi(4) \approx 49.1^\circ$, and it is interesting to investigate further. For \mathbb{R}^N , π_{123}^{Ns} is the sp for the standard axes spacing $\mathbf{d}^0 = (0, 1, 2, \dots, N-2, N-1)$. The sp , after the translation of \bar{X}_1 to position $\bar{X}_{1'}$ with axis spacing $\mathbf{d}^1 = (N, 1, 2, \dots, N-2, N-1)$, is π_{123}^{Ns} . The equations derived from (5.51) are:

$$\begin{cases} \pi_{123}^{Ns} : x_1 - 2x_2 + x_3 = 0, \\ \pi_{1'23}^{Ns} : x_1 + (N-2)x_2 + (1-N)x_3 = 0. \end{cases} \quad (5.61)$$

The angle function is

$$\phi(N) = \cos^{-1} \left(\frac{\sqrt{3(N-2)}}{2\sqrt{3-3N+N^2}} \right), \quad (5.62)$$

some of whose values are $\phi(5) \approx 47.88^\circ$, $\phi(6) = 45^\circ$, and the $\lim_{N \rightarrow \infty} \phi(N) = 30^\circ$.

Next let us compute the 1-flat (line) intersection $\ell_\pi = \pi^{4s} \cap \pi$ where the plane $\pi \subset \mathbb{R}^4$ to find the index-points representation for π , first for a general axis spacing $\mathbf{d} = (d_1, d_2, d_3, d_4)$ described by

$$\begin{cases} \pi : c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 = c_0, \\ \pi_{123}^{4s} : (d_3 - d_2)x_1 + (d_1 - d_3)x_2 + (d_2 - d_1)x_3 = 0, \\ \pi_{234}^{4s} : (d_4 - d_3)x_2 + (d_2 - d_4)x_3 + (d_3 - d_2)x_4 = 0. \end{cases} \quad (5.63)$$

With the notation

$$\begin{cases} a = c_1(d_2 - d_1) + c_3(d_2 - d_3) + c_4(d_2 - d_4), \\ b = c_2(d_2 - d_1) + c_3(d_3 - d_1) + c_4(d_4 - d_1), \end{cases} \quad (5.64)$$

the line intersection is given by

$$\ell_\pi : x_2 = -\frac{a}{b}x_1 + \frac{(d_2 - d_1)}{b}c_0, \quad (5.65)$$

since $\ell_\pi \subset \pi_{1234}^{4s}$, $\bar{\ell}_{\pi_{ij}} = \bar{\ell}_{\pi_{kr}}$, for distinct indices $i, j, k, r \in (1, 2, 3, 4)$, so that in homogeneous coordinates, we obtain

$$\begin{aligned} \bar{\ell}_{\pi_{12}} &= \left(\left((d_2 - d_1)b + d_1 \sum_{i=1}^4 c_i, (d_2 - d_1)c_0, \sum_{i=1}^4 c_i \right) \right) \\ &= \left(\sum_{i=1}^4 c_i d_i, c_0, \sum_{i=1}^4 c_i \right). \end{aligned} \quad (5.66)$$

Substituting the axis spacing \mathbf{d}_N^i in the above, the indexed points for the representation of π are obtained below, where $S = \sum_{i=1}^4 c_i$:

$$\left\{ \begin{array}{l} \bar{\pi}_{0'} = \bar{\pi}_{1234} = (c_2 + 2c_3 + 3c_4, c_0, S), \\ \bar{\pi}_{1'} = \bar{\pi}_{1'234} = (4c_1 + c_2 + 2c_3 + 3c_4, c_0, S) \\ \bar{\pi}_{2'} = \bar{\pi}_{1'2'34} = (4c_1 + 5c_2 + 2c_3 + 3c_4, c_0, S), \\ \bar{\pi}_{3'} = \bar{\pi}_{1'2'3'4} = (4c_1 + 5c_2 + 6c_3 + 3c_4, c_0, S), \\ \bar{\pi}_{4'} = \bar{\pi}_{1'2'3'4'} = (4c_1 + 5c_2 + 6c_3 + 7c_4, c_0, S). \end{array} \right. \quad (5.67)$$

We wrote this out in full detail to show that again, the distance between adjacently indexed points, appearing on the left above in simplified notation, is as for \mathbb{R}^3 proportional (equal to the dimension) to the corresponding coefficient. Specifically,

$$\left\{ \begin{array}{l} \bar{\pi}_{1'} - \bar{\pi}_{0'} = (4c_1, 0, 0), \\ \bar{\pi}_{2'} - \bar{\pi}_{1'} = (4c_2, 0, 0), \\ \bar{\pi}_{3'} - \bar{\pi}_{2'} = (4c_3, 0, 0), \\ \bar{\pi}_{4'} - \bar{\pi}_{3'} = (4c_4, 0, 0). \end{array} \right. \quad (5.68)$$

From this the relations analogous to (8.60) are immediately found, i.e.,

$$x(\bar{\pi}_{3'}) = 12 - [x(\bar{\pi}_{0'}) + x(\bar{\pi}_{1'}) + x(\bar{\pi}_{2'})], \quad x(\bar{\pi}_{4'}) = 4 + x(\bar{\pi}_{0'}). \quad (5.69)$$

To reduce the cumbersome notation, the sp equations (5.56), (5.57), (5.58), (5.59), (5.60) are referred to by $\pi_{0'}^{4s}, \pi_{1'}^{4s}, \pi_{2'}^{4s}, \pi_{3'}^{4s}, \pi_{4'}^{4s}$ respectively. Next we show that the indexed points can be easily obtained in general and the property in (5.68) holds for all N .

Exercises

1. State explicitly the equations for the first sp in \mathbb{R}^5 .
2. Perform the standard translations of the five coordinate axes in \mathbb{R}^5 to obtain explicitly the corresponding rotated five sp .
3. Obtain the coordinates of the six points arising in the representation of a 4-flat in \mathbb{R}^5 . This is the generalization of (5.67).

5.4.2 Indexed Points in \mathbb{R}^N

Remarkably, the collinearity property (as in Theorem 5.2.1) generalizes to higher dimensions, enabling the recursive (on the dimensionality) construction of the representation of p -flats for $2 \leq p \leq N-1$. To achieve this, some intermediate steps are needed. The indexed point corresponding to the axis spacing \mathbf{d}_N^i (i.e., obtained

from the translation of the axes $\bar{X}_1, \dots, \bar{X}_i$, to the positions $\bar{X}_{1'}, \dots, \bar{X}_{i'}$, see (9.1), is denoted by $\bar{\pi}_{i'}$.

Theorem 5.4.1 (B. Dimsdale) *The 1-fla $\pi \cap \pi_{i'}^{Ns}$, where $i = 0, 1, \dots, N$ and*

$$\pi : \sum_{k=1}^N c_k x_k = c_0, \quad (5.70)$$

is a hyperplane in \mathbb{R}^N ; an $(N-1)$ -fla is represented by the point

$$\bar{\pi}_{i'} = \left(\sum_{k=1}^N d_{ik} c_k, c_0, \sum_{k=1}^N c_k \right), \quad (5.71)$$

where the d_{ik} are the interaxis distances for the spacing \mathbf{d}_N^i as given in (9.1). Explicitly, using (9.3),

$$\bar{\pi}_{i'} = \left(\sum_{k=1}^N (k-1 + N S_i(k)) c_k, c_0, \sum_{k=1}^N c_k \right). \quad (5.72)$$

Proof. To determine the 1-flat and obtain its representation, it suffices to find two points in the intersection $\pi \cap \pi_{i'}^{Ns}$.

Step 1. Points $P_j = (p_{1j}, p_{2j}, \dots, p_{Nj}) \in \pi_{i'}^{Ns}$ are such that

$$\exists m_j, b_j \in \mathbb{R} \ni p_{kj} = m_j d_k + b_j. \quad (5.73)$$

Step 2. So for $P_j \in \pi \cap \pi_{i'}^{Ns}$, in addition to (5.73), the coordinates of P_j must also satisfy (5.70), i.e.,

$$\sum_{k=1}^N c_k (m_j d_k + b_j) = c_0.$$

Step 3. By Step 1, \bar{P}_j is the straight line $y = m_j x + b_j$ in the xy plane, and together with Step 2 we have

$$\sum_{k=1}^N c_k \{m_j d_k + (y - m_j x)\} = c_0,$$

or

$$-m_j \left(\sum_{k=1}^N c_k \right) x + \left(\sum_{k=1}^N c_k \right) y = -m_j \left(\sum_{k=1}^N d_k c_k \right) + c_0; \quad (5.74)$$

this is an alternative form of the equation for the line \bar{P}_j .

Step 4. Since xy is the projective plane \mathbb{P}^2 , for *any* two distinct points P_1, P_2 the corresponding lines \bar{P}_1 and \bar{P}_2 intersect at

$$\begin{aligned} (-m_1 + m_2) \left(\sum_{k=1}^N c_k \right) x &= (-m_1 + m_2) \left(\sum_{k=1}^N d_k c_k \right) \\ \Rightarrow \left(\sum_{k=1}^N c_k \right) x &= \left(\sum_{k=1}^N d_k c_k \right). \end{aligned}$$

Substitution in (5.74) provides the remaining coordinate:

$$\left(\sum_{k=1}^N c_k \right) y = c_0.$$

The point (x, y) is independent of the particular lines \bar{P}_1 and \bar{P}_2 used, so all the lines \bar{P}_j for the \bar{P}_j in Step 2 must intersect at the same (x, y) . Converting to homogeneous coordinates yields (5.71). ■

Corollary 5.4.2 (Hyperplane Representation; J. Eickemeyer) *The hyperplane π given by (5.70) is represented by the $N - 1$ points $\bar{\pi}_{i'}$ with N indices given by (5.72), for $i = 0, 1, 2, \dots, (N - 2)$.*

Specifically,

$$\left\{ \begin{array}{l} \bar{\pi}_{0'} = \bar{\pi}_{12\dots N} = (c_2 + 2c_3 + \dots + (k-1)c_k + \dots + (N-1)c_N, c_0, S), \\ \bar{\pi}_{1'} = (Nc_1 + c_2 + \dots + (k-1)c_k + \dots + (N-1)c_N, c_0, S), \\ \bar{\pi}_{2'} = (Nc_1 + (N+1)c_2 + \dots + (k-1)c_k + \dots + (N-1)c_N, c_0, S), \\ \dots \\ \bar{\pi}_{i'} = (Nc_1 + \dots + (N+i-1)c_i + ic_{i+1} + \dots + (N-1)c_N, c_0, S), \quad i \geq 1, \\ \dots \\ \bar{\pi}_{(N-2)'} = (Nc_1 + \dots + (2N-3)c_{N-2} + (N-2)c_{N-1} + (N-1)c_N, c_0, S), \\ \bar{\pi}_{(N-1)'} = (Nc_1 + \dots + (2N-3)c_{N-2} + (2N-2)c_{N-1} + (N-1)c_N, c_0, S), \\ \bar{\pi}_{N'} = (Nc_1 + \dots + (2N-3)c_{N-2} + (2N-2)c_{N-1} + (2N-1)c_N, c_0, S), \end{array} \right. \quad (5.75)$$

where $S = \sum_{k=1}^N c_k$. The first $N - 1$ points suffice for the representation. As was done for \mathbb{R}^3 and \mathbb{R}^4 , it is useful to generate the two additional points $\bar{\pi}'_{N-1}, \bar{\pi}'_N$,

which here, based on the axis spacing $\mathbf{d}_N^{N-1}, \mathbf{d}_N^N$, with the their x coordinates (analogous to (5.69)) given by

$$x_{(N-1)'} = N(N-1) - \sum_{i=0}^{N-2} x_{i'}, \quad x'_N = N + x'_0, \quad (5.76)$$

with the compact notation $x_{i'} = x(\bar{\pi}_{i'})$.

A p -flat in \mathbb{R}^N is specified by $N - p$ linearly independent linear equations, which, without loss of generality, can be of the form

$$\pi^p : \begin{cases} \pi_{12\dots(p+1)}^p : c_{11}x_1 + \dots + c_{(p-1)1}x_p + c_{p1}x_{p+1} = c_{10}, \\ \pi_{23\dots(p+2)}^p : c_{22}x_2 + \dots + c_{p2}x_{p+1} + c_{(p+1)2}x_{p+2} = c_{20}, \\ \dots \\ \pi_{j\dots(p+j)}^p : c_{jj}x_j + \dots + c_{(p+j-1)j}x_{p+j-1} + c_{(p+j)j}x_{p+j} = c_{j0}, \\ \dots \\ \pi_{(N-p)\dots N}^p : c_{(N-p)(N-p)}x_{N-p} + \dots \\ \quad + c_{(N-1)(N-p)}x_{N-1} + c_{N(N-p)}x_N = c_{(N-p)0}, \end{cases}$$

and is rewritten compactly as

$$\pi^p : \left\{ \pi_{j\dots(p+j)}^p : \sum_{k=i}^{p+j} c_{jk}x_k = c_{j0}, j = 1, 2, \dots, (N-p) \right\}. \quad (5.77)$$

A p -flat $\pi^p \subset \mathbb{R}^N$ is the intersection of $N - p$ hyperplanes, and (5.4.2) is the analogue of the “adjacent-variable” description for lines in Chapter 4 with analogous indexing. Unless otherwise specified, a p -flat is described by (5.4.2) with the standard spacing \mathbf{d}_N^0 .

Theorem 5.4.3 (J. Eickemeyer) *A p -flat in \mathbb{R}^N given by (5.77) is represented by the $(N - p)p$ points with $p + 1$ indices*

$$\bar{\pi}_{\{j\dots(p+j)\}_{i'}}^p = \left(\sum_{k=1}^{p+1} d_{ik}c_{jk}, c_{j0}, \sum_{k=1}^{p+1} c_{jk} \right), \quad (5.78)$$

where $j = 1, 2, \dots, N - p, i = 1, 2, \dots, p$, and the d_{ik} are the distances specified by the axis spacing \mathbf{d}_N^i .

Proof. Each $\pi_{j\dots(p+j)}^p$ in (5.77) can be considered as a hyperplane in $\mathbb{R}^{(p+1)}$: $x_j \dots x_{j+p+1}$, whose representation, according to Corollary 5.4.2, consists of the p points $\bar{\pi}_{\{j\dots(p+j)\}_{i'}}^p, i = 1, \dots, p$. They are prescribed by the axis spacing $\mathbf{d}_N^i = (d_{i1}, \dots, d_{i(p+1)}, \dots, N)$ as Theorem 5.4.1. There are $(N - p)$ hyperplanes described by (5.77); therefore there are $(N - p)p$ such points altogether. ■

♣ FT-6 To clarify, a hyperplane π in \mathbb{R}^4 π (i.e., 3-flat) can be represented by the three points

$$\bar{\pi}_{1234}, \bar{\pi}_{1'234}, \bar{\pi}_{1'2'34}, \quad (5.79)$$

while for a 2-flat π^2 , $p = 2$, $N = 4$, $p(N - p) = 4$, and it is represented by the four points

$$\pi_{123}^2 : \bar{\pi}_{123}^2, \bar{\pi}_{1'23}^2; \quad \pi_{234}^2 : \bar{\pi}_{234}^2, \bar{\pi}_{2'34}^2. \quad (5.80)$$

Similarly in \mathbb{R}^5 , a hyperplane π is represented by four points, a 3-flat π^3 and a 2-flat π^2 by six points each as, indicated below:

$$\left\{ \begin{array}{l} \pi : \bar{\pi}_{12345}, \bar{\pi}_{1'2345}, \bar{\pi}_{1'2'345}, \bar{\pi}_{1'2'3'45}, \\ \pi^3 : \left\{ \begin{array}{l} \pi_{1234}^3 : \bar{\pi}_{1234}^3, \bar{\pi}_{1'234}^3, \bar{\pi}_{1'2'34}^3, \\ \pi_{2345}^3 : \bar{\pi}_{2345}^3, \bar{\pi}_{2'345}^3, \bar{\pi}_{2'3'45}^3, \end{array} \right. \\ \pi^2 : \left\{ \begin{array}{l} \pi_{123}^2 : \bar{\pi}_{123}^2, \bar{\pi}_{1'23}^2, \\ \pi_{234}^2 : \bar{\pi}_{234}^2, \bar{\pi}_{2'34}^2, \\ \pi_{345}^2 : \bar{\pi}_{345}^2, \bar{\pi}_{3'45}^2. \end{array} \right. \end{array} \right. \quad (5.81)$$

In many instances it is possible to use simplified notation by just retaining the subscripts so that the three points in (5.79) are referred to by 1234, 1'234, 1'2'34. The dimensionality is one less than the number of indices. Continuing, the points representing π^3 in (5.81) are denoted simply by 1234, 1'234, 1'2'34; 2345, 2'345, 2'3'45. Since there two sets of three points with five different indices altogether, we can conclude that this is a 2-flat in \mathbb{R}^5 . The simplified and the more formal notation are used interchangeably as in Section 5.5. This theorem unifies all previous results for p -flats π^p with $0 \leq p < N$.

♣ FT-6e

5.4.3 Collinearity Property

The underpinning of the construction algorithm for the point representation of a 2-flat $\pi^2 \subset \mathbb{R}^3$, as we saw, is the collinearity property. Namely, for *any* $\pi^1 \subset \pi^2$, the points $\bar{\pi}_{12}^1, \bar{\pi}_{13}^1, \bar{\pi}_{23}^1$ are collinear with $\bar{\pi}_{123}$. For the generalization to p -flats, let

$$\bar{L}_j^{P_k} = \bar{\pi}_{j \dots (p+j)}^{P_k} \bullet \bar{\pi}_{(j+1) \dots (p+j+1)}^{P_k} \quad (5.82)$$

denote the line $\bar{L}_j^{P_k}$ on the indicated two points. The topic of this section is the proof that $\pi^{(p-1)_1}, \pi^{(p-1)_2} \subset \pi^p \subset \mathbb{R}^N$,

$$\bar{\pi}_{j \dots (p+j+1)}^p = \bar{L}_j^{(p-1)_1} \cap \bar{L}_j^{(p-1)_2}. \quad (5.83)$$

As an example, for $j = 1, p = 2, N = 3$, our old friend from Section 5.2.2 is recast as

$$\bar{L}_1^{\pi^{1_k}} = \bar{\pi}_{12}^{1_k} \bullet \bar{\pi}_{23}^{1_k}, \quad k = 1, 2, \quad \bar{\pi}_{123}^2 = \bar{L}_1^{\pi^{1_1}} \cap \bar{L}_1^{\pi^{1_2}}.$$

The pair (5.82) and (5.83) state the basic recursive construction implied in the *representation mapping* stated formally below. The recursion is on the dimensionality, increased by one at each stage, of the flat whose representative points are constructed. Though the notation may seem intimidating, the idea is straightforward, and to clarify it we illustrate it for a hyperplane $\pi^3 \subset \mathbb{R}^4$ in Figs. 5.44 and 5.46 starting from the four points $\pi^{0_1}, \pi^{0_2}, \pi^{0_3}, \pi^{0_4}$ with \rightarrow pointing to the construction result. Diagrammatically, the sequence of steps is

$$\left. \begin{array}{l} \pi^{0_1}, \pi^{0_2} \rightarrow \pi^{1_1} \\ \pi^{0_2}, \pi^{0_3} \rightarrow \pi^{1_2} \end{array} \right\} \rightarrow \pi^{2_1} (\bar{\pi}_{123}^{2_1}, \bar{\pi}_{1'23}^{2_1}) \left. \begin{array}{l} \pi^{0_2}, \pi^{0_3} \rightarrow \pi^{1_2} \\ \pi^{0_3}, \pi^{0_4} \rightarrow \pi^{1_3} \end{array} \right\} \rightarrow \pi^{2_2} (\bar{\pi}_{234}^{2_2}, \bar{\pi}_{2'34}^{2_2}) \left. \begin{array}{l} \pi^{0_1}, \pi^{0_2} \rightarrow \pi^{1_1} \\ \pi^{0_2}, \pi^{0_3} \rightarrow \pi^{1_2} \\ \pi^{0_3}, \pi^{0_4} \rightarrow \pi^{1_3} \end{array} \right\} \rightarrow \pi^3 (\bar{\pi}_{1234}^3, \bar{\pi}_{1'234}^3, \bar{\pi}_{1'2'34}^3). \quad (5.84)$$

For the construction of a regular (i.e., not ideal) flat, the flats $\pi^p \subset \pi^3$ with dimensionality $3 > p \geq 1$ in the construction must have a nonempty intersection. From the polygonal lines representing $\pi^{0_1}, \pi^{0_2}, \pi^{0_3}$, the two 1-flats π^{1_1}, π^{1_2} are constructed with $\pi^{1_1} \cap \pi^{1_2} = \pi^{0_2}$, yielding the points $\bar{\pi}_{12}^{1_1}, \bar{\pi}_{23}^{1_1}, \bar{\pi}_{34}^{1_1}$, as shown in Fig. 5.44 (left). The portion of the construction involving x_1, x_2, x_3 is shown in Cartesian coordinates in Fig. 5.45. The three representing points for each 1-flat are joined by *two* lines to form polygonal lines having three vertices (the

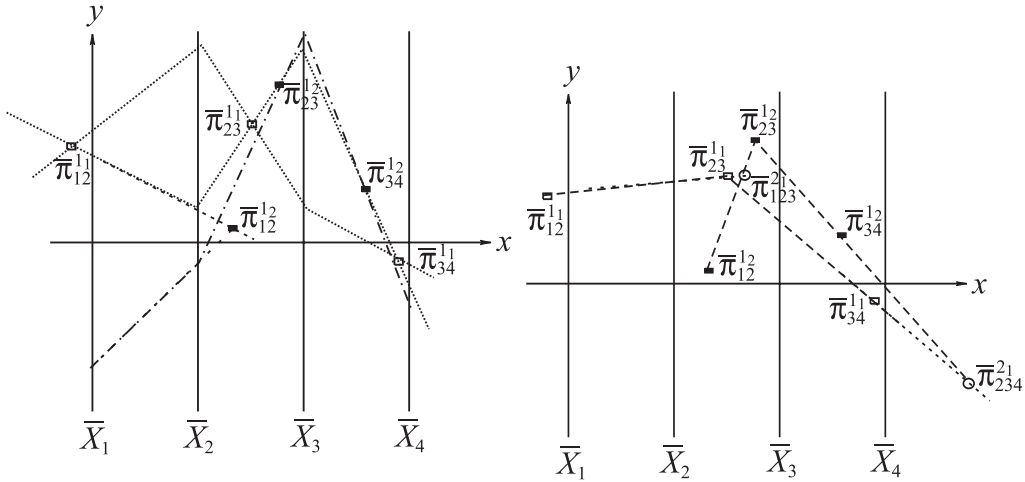


Figure 5.44. ♣ **FT-7** Recursive construction in \mathbb{R}^4 .

A pair of points π^{0_1}, π^{0_2} determines a line (1-flat) π^{1_1} represented by the three constructed points $\bar{\pi}^{1_1}_{i-1,i}, i = 2, 3, 4$ (left). Another 1-flat π^{1_2} is determined by one of these points π^{0_2} and an additional point π^{0_3} as represented by the three black points. Since $\pi^{1_1} \cap \pi^{1_2} = \pi^{0_2}$, the two 1-flats determine a 2-flat π^{2_1} , and two of its representing points $\bar{\pi}^{2_1}_{123}, \bar{\pi}^{2_1}_{234}$ are seen on the right. They are the intersections of the two polygonal lines joining the points previously obtained representing the two 1-flats.

points). From the intersection of these new polygonal lines the points $\bar{\pi}^{2_1}_{123}, \bar{\pi}^{2_1}_{234}$, representing a 2-flat contained on π^3 , are constructed as shown on the right of Fig. 5.44. Similarly, π^{2_2} is constructed from the three points $\pi^{0_2}, \pi^{0_3}, \pi^{0_4}$ in the same way.

At any stage, a point representing $\bar{\pi}^r$, where the superscript is the flat's dimension, is obtained by the intersection of *any pair* of lines joining points representing flats of dimension $r - 1$ contained in π^r .

The axes \bar{X}_1, \bar{X}_2 are each translated four units to the right, and construction proceeds until all three representing points $\bar{\pi}^3_{1234}, \bar{\pi}^3_{1'234}, \bar{\pi}^3_{1'2'34}$ are obtained. The first translation is shown in Fig. 5.46 on the right. As a reminder, these points represent the *lines* that are the intersections of π^3 with the first superplane, due to the standard spacing with $d_1 = 0, d_2 = 1, d_3 = 2, d_4 = 3$, followed by the second sp with $d'_1 = 4$ and then the third sp with $d'_2 = 5$ in \mathbb{R}^4 .

Theorem 5.4.4 (Collinearity Construction Algorithm) For any $\pi^{(p-2)} \subset \pi^{(p-1)} \subset \mathbb{R}^N$, the points $\bar{\pi}^{(p-2)}_{1...(p-1)}, \bar{\pi}^{(p-2)}_{2...(p-1)p}, \bar{\pi}^{(p-1)}_{1...(p-1)p}$ are collinear.

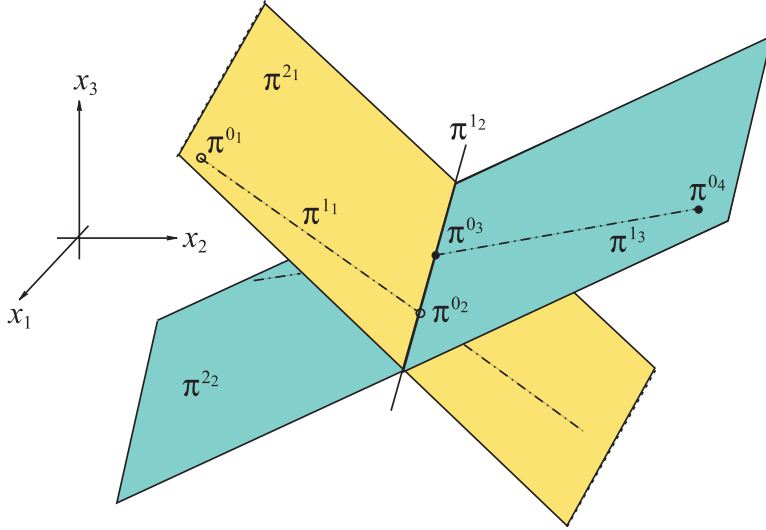


Figure 5.45. The construction of the $x_1x_2x_3$ part of the 3-flat π^3 from four points (0-flats).

Proof. Step 1. Let the $(p-1)$ - and $(p-2)$ -flats be given by

$$\begin{cases} \pi^{(p-1)} : \pi_{i\dots(p+i-1)}^{(p-1)} : \sum_{k=i}^{p+i-1} c_{ik}x_k = c_{0i}, & i = 1, \dots, N-p+1, \\ \pi^{(p-2)} : \pi_{j\dots(p+j-2)}^{(p-2)} : \sum_{k=j}^{p+j-2} a_{jk}x_k = a_{0j}, & j = 1, \dots, N-p+2. \end{cases} \quad (5.85)$$

Step 2. Consider two distinct points $A^r = (\alpha_1^r, \dots, \alpha_N^r)$, $r = 1, 2 \in \pi^{(p-2)}$, and substitute their first p -components in the equation for $\pi_{12\dots p}^{(p-1)}$ in (5.85) to obtain

$$c_{11}\alpha_1^r + \dots + c_{(p-1)1}\alpha_{(p-1)}^r + c_{p1}\alpha_p^r = c_{01}. \quad (5.86)$$

Substitution in the first two equations for $\pi^{(p-2)}$, i.e., $\pi_{12\dots(p-2)}^{(p-2)}$, $\pi_{2\dots(p-2)(p-1)}^{(p-2)}$, yields

$$\begin{cases} \pi_{12\dots(p-2)}^{(p-2)} : a_{11}\alpha_1^r + \dots + a_{(p-2)1}\alpha_{(p-2)}^r + a_{(p-1)1}\alpha_{(p-1)}^r = a_{01}, \\ \pi_{2\dots(p-2)(p-1)}^{(p-2)} : a_{22}\alpha_2^r + \dots + a_{(p-1)2}\alpha_{(p-1)}^r + a_{p2}\alpha_p^r = a_{02}, \end{cases}$$

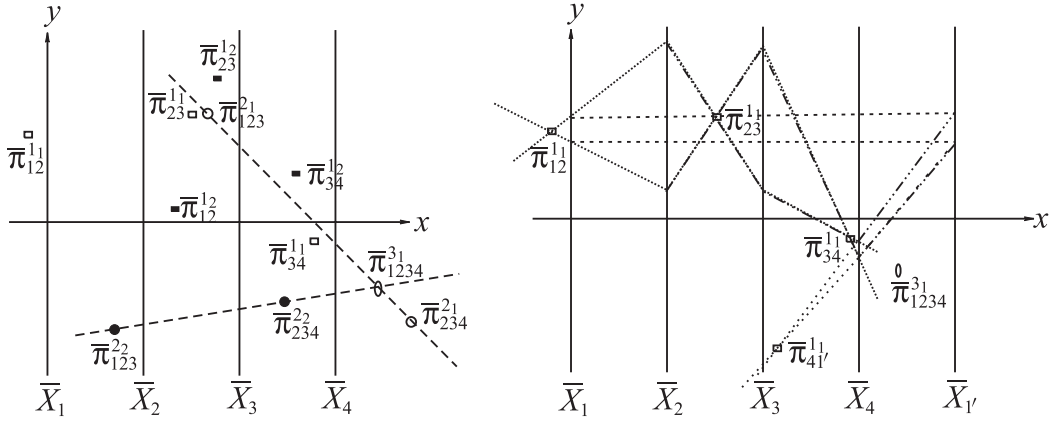


Figure 5.46. Recursive construction in \mathbb{R}^4 continued.

Two 2-flats π^{21} , previously constructed, and another π^{22} represented by the two black points (left), determine a 3-flat π^3 . Pairs of points representing the same 2-flat are joined, and their intersection is the point $\bar{\pi}_{1234}^3$. This is one of the three points representing the 3-flat. The “debris” from the previous constructions, points with fewer than four indices, can now be discarded. A new axis (right) $\bar{X}_{1'}$ is placed one unit to the right of \bar{X}_3 , and the x_1 values are transferred to it from the \bar{X}_1 axis. Points are now represented by new polygonal lines between the \bar{X}_2 and $\bar{X}_{1'}$ axes, and one of the points $\bar{\pi}_{41'}^1$, representing the 1-flat π^{11} on the new triple of \parallel -coordinate axes, is constructed as in the first step. ♣ FT-7e

whose sum is

$$\begin{aligned} & a_{11}\alpha_1^r + (a_{21} + a_{22})\alpha_2^r + \cdots + (a_{(p-1)1} + a_{(p-1)2})\alpha_{(p-1)}^r + a_{p2}\alpha_p^r \\ &= a_{01} + a_{02}. \end{aligned} \quad (5.87)$$

Step 3. Equations (5.86) and (5.87) are the same, since subtracting one from the other yields

$$\begin{aligned} & \alpha_1^r(c_{11} - a_{11}) + \alpha_2^r\{c_{21} - (a_{21} + a_{22})\} + \cdots + \alpha_p^r(c_{p1} - a_{p2}) \\ &= c_{01} - (a_{01} + a_{02}). \end{aligned} \quad (5.88)$$

Letting $r = 1, 2$ successively in (5.88) provides two equations whose difference is

$$[\alpha_1^1 - \alpha_1^2]b_1 + [\alpha_2^1 - \alpha_2^2]b_2 + \cdots + [\alpha_p^1 - \alpha_p^2]b_p = 0, \quad (5.89)$$

the b_i being the coefficients of (5.88). This is effectively an *identity* for every pair of distinct points in $\pi^{(p-2)}$. Hence the coefficients and hence the right-hand-side of (5.88) must vanish:

$$\begin{aligned}
c_{01} &= (a_{01} + a_{02}), \quad c_{11} = a_{11}, \quad c_{p1} = a_{p2}, \quad c_{k1} = (a_{k1} + a_{k2}), \\
k &= 2, \dots, p-1.
\end{aligned} \tag{5.90}$$

Step 4. The homogeneous coordinates of the three points in question as obtained from Theorem 5.4.3 are

$$\begin{aligned}
\bar{\pi}_{1\dots(p-1)}^{(p-2)} &= \left(\sum_{k=1}^{p-1} d_{1k} a_{1k}, a_{01}, \sum_{k=1}^{p-1} a_{1k} \right), \\
\bar{\pi}_{2\dots(p-1)p}^{(p-2)} &= \left(\sum_{k=2}^p d_{2k} a_{2k}, a_{02}, \sum_{k=2}^p a_{2k} \right),
\end{aligned}$$

and

$$\bar{\pi}_{12\dots(p-1)p}^{(p-1)} = \left(\sum_{k=1}^p d_{1k} c_{2k}, c_{01}, \sum_{k=1}^p c_{1k} \right).$$

Note that for this portion of the construction, the axis spacing, is $d_{1k} = d_{2k}$, $k = 1, 2, \dots, p$. Forming the determinant from these homogeneous coordinates yields

$$\begin{vmatrix}
\sum_{k=1}^{p-1} d_{1k} a_{1k} & a_{01} & \sum_{k=1}^{p-1} a_{1k} \\
\sum_{k=2}^p d_{2k} a_{2k} & a_{02} & \sum_{k=2}^p a_{2k} \\
\sum_{k=1}^p d_{1k} c_{2k} & c_{01} & \sum_{k=1}^p c_{1k}
\end{vmatrix}.$$

From Step 3 and the observation on the axis spacing, the last row is the sum of the first and second rows; thus the value of the determinant is zero, showing that the three points are collinear. ■

Corollary 5.4.5. *For any $\pi^{(p-2)} \subset \pi^{(p-1)} \subset \mathbb{R}^N$, the points $\bar{\pi}_{\{j\dots(p+j-2)\}_{i'}}^{(p-2)}$, $\bar{\pi}_{\{(j+1)\dots(p+j-1)\}_{i'}}^{(p-2)}$, $\bar{\pi}_{\{j\dots(p+j-1)\}_{i'}}^{(p-1)}$ are collinear.*

The proof is the same, taking proper care to keep track of the indices and the corresponding axis spacing. The recursive construction is illustrated for a 5-flat in \mathbb{R}^6 from Figs. 5.47 through 5.49.

Exercises

1. What points represent the sp in \mathbb{R}^N ?

2. State the coordinates of the five points representing a 5-flat $\pi \subset \mathbb{R}^6$.
3. State the points representing a 4-flat π^4 , 3-flat π^3 , and 2-flat π^2 in \mathbb{R}^6 as in (5.81).
4. Prove (5.76) for \mathbb{R}^N .

5.5 ** Construction Algorithms in \mathbb{R}^4

The generalization of the 3-D construction algorithms to 4-D is direct and is an opportune time to introduce simplified notation, which is used when the context is clear.

5.5.1 The Five Indexed Points

The first example is the equivalent of the four-indexed-point algorithm in Section 5.3.2 for 4-D. That is, given three points $\bar{\pi}_{1234}$, $\bar{\pi}_{1'234}$, $\bar{\pi}_{1'2'34}$ specifying a 3-flat $\pi^3 : c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 = c_0$ in \mathbb{P}^4 , we wish to construct the other two, $\bar{\pi}_{1'2'3'4}$ and $\bar{\pi}_{1'2'3'4'}$. We start by revisiting the last step in the recursive construction Fig. 5.46 and show it again in Fig. 5.50 (left). This is the stage where two lines \bar{P}_1 , \bar{P}_2 each on the 123 and 234 points of a 2-flat determine the first point $\bar{\pi}_{1234}$ representing a 3-flat π^3 , a situation completely analogous to that in Fig. 5.10, where here the dimensionality of the objects involved is raised by one. Of course, the lines represent points $P_1 \in \pi^4s_1 \cap \pi^{21}$ and $P_2 \in \pi^4s_1 \cap \pi^{22}$ for two 2-flats π^{21} , π^{22} contained in π^3 . As a reminder, containment is seen by the *on* relation. For example, line \bar{P} is *on* point $\bar{\pi}_{123}^2 \Leftrightarrow P \in \pi^2$, and here in particular the point P is on the line $\pi_{123}^2 = \pi^2 \cap \pi_1^s$. As for the 3-D case, the point $\bar{\pi}_{1234}$ represents a line on an *sp*. Specifically, $\bar{\pi}_{1234} = \pi_1^{4s} \cap \pi^3$. From a line \bar{P}_1 on the points $\bar{\pi}_{2341}$ (which we assume has been previously constructed) and $\bar{\pi}_{1'231}$, the second point $\bar{\pi}_{1'234}$ is determined, as shown on the right of Fig. 5.50.

The algorithm's input consists of three points $\bar{\pi}_{1234}^3$, $\bar{\pi}_{1'234}^3$, $\bar{\pi}_{1'2'34}^3$, shown in simplified notation 1234, 1'234, 1'2'34 on the left of Fig. 5.51. Four points determine a hyperplane $\pi^3 \subset \mathbb{R}^4$. On the right, three lines \bar{P} , \bar{P}' , \bar{P}'' are chosen on the points 1234, 1'234, 1'2'34 respectively. Clearly $P \in \pi^3 \cap \pi_1^{4s}$, $P' \in \pi^3 \cap \pi_{1'}^{4s}$, $P'' \in \pi^3 \cap \pi_{1''}^{4s}$, and these three points determine a 2-flat $\pi^2 \subset \pi^3$, which can be described by

$$\pi^2 : \begin{cases} \pi_{123}^2 : a_1x_1 + a_2x_2 + a_3x_3 = a_0, \\ \pi_{234}^2 : b_2x_2 + b_3x_3 + b_4x_4 = b_0. \end{cases} \quad (5.91)$$

As for a 1-flat in a 2-flat, see Fig. 5.20; also, for a 2-flat in a 3-flat $\bar{P} \cap \bar{P}'$, $\bar{\pi}_{234}^2$ is one of the points representing π^2 . It is denoted by **234** and is on the horizontal line **234**, so that $2'34$ is **234** $\cap \bar{P}''$. These two points represent π_{234}^2 given by (5.91), a 2-flat in the $x_2x_3x_4$ subspace of \mathbb{R}^4 . The points $\bar{\pi}_{2'3'4}^2$, $\bar{\pi}_{2'3'4'}^2$, denoted by $2'3'4$, $2'3'4'$, are needed next and are determined from **234**, $2'34$ via the construction algorithm in Section 5.3.2. For the representation of the 2-flat π_{123}^2 as in (5.91), *any* point $\bar{\pi}_{123}^2$ can be chosen on \bar{P} , such as the one denoted by **123** in Fig. 5.52. The intersections of the horizontal line **123** on **123** and \bar{P}' , \bar{P}'' determine the points $1'23$, $1'2'3$. Four points determine a hyperplane $\pi^3 \subset \mathbb{R}^4$, and so far we have used three. Proceeding as shown in Fig. 5.52, the line \bar{P}''' is drawn on the point $2'3'4$ and *parallel* to \bar{P} ; the point P''' is the point P in the $\bar{X}_{1'}$, $\bar{X}_{2'}$, $\bar{X}_{3'}$, $\bar{X}_{4'}$ axes, just as for the 3-D case but here $P''' \in \pi^3 \cap \pi_{1'2'3'4'}^{4s}$ with $\pi_{1'2'3'4'}^{4s}$ coinciding with π_{1234}^{4s} . The intersections of \bar{P}''' with the lines **123**, **1234** determine the points $1'2'3'$ and $1'2'3'4$. The construction is completed by drawing the line \bar{P}^{iv} on the points $2'3'4'$, $1'2'3'$ whose intersection with **1234** provides the the fifth point of the representation of π^3 . The four independent points P , P' , P'' , P^{iv} determine π^3 . The algorithm's output is invariant with respect to the choice of points as well as the selection of π^2 (see Exercise 3). Other equivalent determinations for the 3-flat are one point P and a 2-flat π^2 , or three lines, which are the 3-flat's intersection with the *sp* providing the three indexed points' representation.

By the way, it is immediately clear that a 2-flat $\pi^2 \subset \pi^3 \Leftrightarrow \bar{\pi}_{123}^2, \bar{\pi}_{234}^2, \bar{\pi}_{1234}^3$ are collinear and equivalently for the remaining points representing π^3 . This is the generalization of the $\ell \subset \pi^2$ condition shown in Fig. 5.20 and of course a particular instance of the recursion on the dimensionality collinearity construction.

5.5.2 Intersecting Hyperplanes in \mathbb{R}^4

Next we provide an algorithm that constructs the intersection of two 3-flats π^{3_1} , π^{3_2} , the result being a 2-flat π^2 like the one given in (5.91). The algorithm's input is shown on the left of Fig. 5.53, consisting of three points in \parallel -coords each representing the two 3-flats. On the right, the lines \bar{P} , \bar{P}' , \bar{P}'' joining the pairs **1234**, **1'234**, **1'2'34** represent three points that specify the 2-flat $\pi^2 = \pi^{3_1} \cap \pi^{3_2}$ whose representation is constructed in the next steps. As seen, the point $\bar{\pi}_{234}^2 = \bar{P} \cap \bar{P}'$, **234** also establishes the **234** horizontal line with $\bar{\pi}_{2'34}^2 = \bar{P}' \cap \mathbf{234}$ marked by $2'34$. The points $2'3'4$, $2'3'4'$ and, using the previous algorithm, also $1'2'3'4$ are constructed and for easy reference are marked by dotted circles.

The remaining steps are illustrated in Fig. 5.54:

1. Draw the line \bar{P}^{iv} on the point $2'3'4'$ parallel to \bar{P} , as in the 3-D construction, since the point P^{iv} is the point P but in the rotated coordinates designated by $1', 2', 3', 4'$ in the simplified notation. Necessarily, \bar{P}^{iv} is on the $1'2'3'4'$ coordinate system and on the two horizontal lines $1234_i, i = 1, 2$. The point $2'3'4'$ was constructed previously for this purpose. Alternatively, any of the two points $1'2'3'4'$ can be used, but its construction is more laborious.
2. The line \bar{P}''' on the points $2'3'4, 1'2'3'4_1$ is constructed, providing $1'2'3'4_2 = \bar{P}''' \cap 1234_2$ and $1'2'3' = \bar{P}'' \cap \bar{P}^{iv}$, which establishes the **123** horizontal line.
3. The points $123 = \bar{P} \cap 123, 1'23 = \bar{P}' \cap 123, 1'2'3 = \bar{P}'' \cap 123$ are now available. The coincidence of the points 123 and $1'23$ here is due to the second coefficient $a_2 = 0$ of π_{123}^2 , see (5.91), and is not true in general.

Note that π_{123}^2 is a 3-flat in \mathbb{R}^4 parallel to the x_4 axis, and hence its equation has zero coefficient for x_4 . The locations of the points $123, 1'23, 1'2'3, 1'2'3'$, which are now available together with the first equation of (5.67) and equation (5.68) enable us to obtain the values of the π_{123}^2 coefficients. Similarly, the coefficients of the equation of π_{234}^2 , a 3-flat parallel to the x_1 axis (hence zero coefficient of x_1 is zero), are found. The two equations describe explicitly the algorithm's output π^2 .

Further higher-dimensional constructions are not presented. They offer wonderful opportunities for exercises (below), projects, and research topics.

Exercises

1. Simplify the above constructions using the relations in (5.69).
2. Given four points each in a successive sp , provide an algorithm to construct the 3-flat containing them.
3. For the algorithm in Section 5.5.1, show that the result is invariant for *any* choice of points $P \in \pi^3 \cap \pi_1^{4s}, P' \in \pi^3 \cap \pi_{1'}^{4s}, P'' \in \pi^3 \cap \pi_{1''}^{4s}, P^{iv} \in \pi^3 \cap \pi_{1'2'3'4'}^{4s}$, and π_{123}^2 .
4. Generalize the algorithm in Section 5.3.6 for a point P and 3-flat π^3 in \mathbb{R}^4 .
5. For a given 3-flat π^3 and a 1-flat (line) ℓ , provide conditions for containment $\ell \subset \pi^3$ and $\ell \cap \pi^3 = \emptyset$. What is $\ell \cap \pi^3$?
6. Generalize the algorithm of Section 5.5.1 to \mathbb{R}^5 and construct the *six* points arising in the representation of a 4-flat given the first four points.
7. Generalize the algorithm of Section 5.5.2 to construct the intersection of two 4-flats in \mathbb{R}^5 .

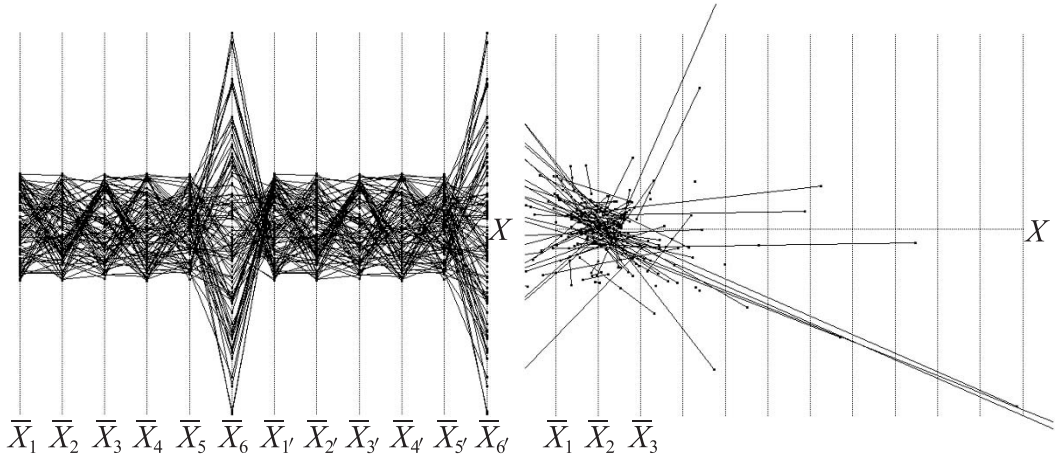


Figure 5.47. Randomly selected points on a hyperplane in \mathbb{R}^6 .

Polygonal lines (left) on the $\bar{X}_1, \dots, \bar{X}_6$ axes representing randomly selected points on a 5-flat $\pi^5 \subset \mathbb{R}^6$. The $\bar{\pi}_{12}^{1_i}, \bar{\pi}_{23}^{1_i}$ portions of the 1-flats $\subset \pi^5$ constructed (right) from the polygonal lines. No pattern is evident.

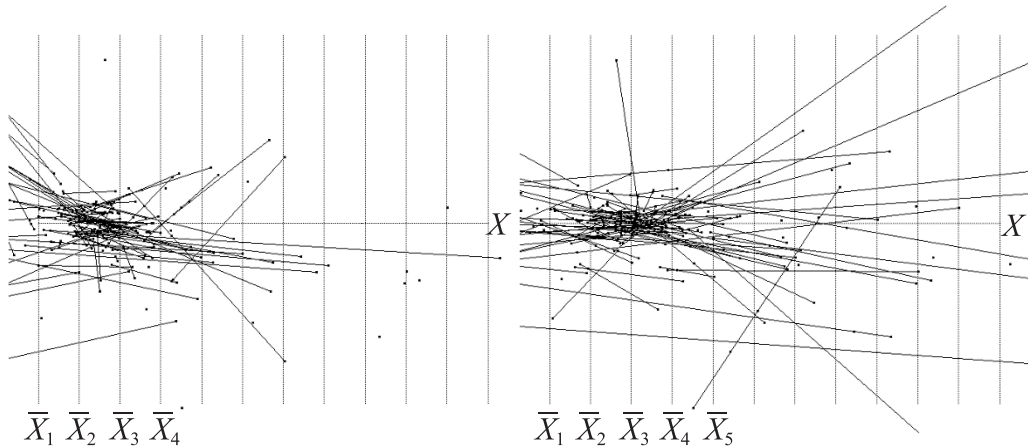


Figure 5.48. The $\bar{\pi}_{123}^{2_i}, \bar{\pi}_{234}^{2_i}$ (left) points for the 2-flats $\subset \pi^5$.

They are constructed from the polygonal lines joining $\bar{\pi}_{12}^{1_i}, \bar{\pi}_{23}^{1_i}, \bar{\pi}_{34}^{1_i}$. The $\bar{\pi}_{123}^{2_i}, \bar{\pi}_{234}^{2_i}$ (right) portions of the 2-flats $\subset \pi^5$ constructed from the polygonal lines joining $\bar{\pi}_{12}^{1_i}, \bar{\pi}_{23}^{1_i}, \bar{\pi}_{34}^{1_i}$.

8. Generalize the rotation of a 2-flat in \mathbb{R}^3 to display in $\|\cdot\|$ -coords the rotation of a 3-flat in \mathbb{R}^4 about an axis (i.e., a 1-flat). If it is not possible to perform this rotation, show *graphically* the constraints involved; **hard**.

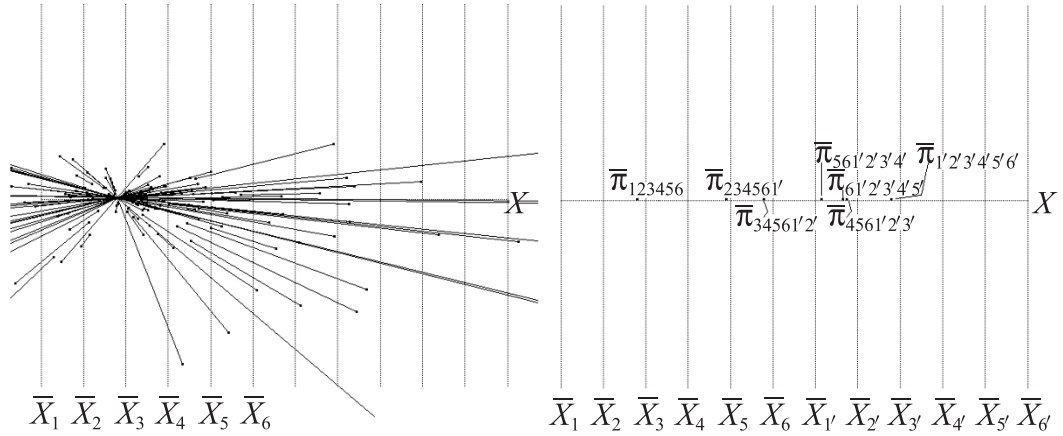


Figure 5.49. This is it!

On the left are the $\bar{\pi}_{12345}^{4i}, \bar{\pi}_{23456}^{4i}$ of the 4-flats $\subset \pi^5$ constructed from the polygonal lines joining $\bar{\pi}_{1234}^{3i}, \bar{\pi}_{2345}^{3i}, \bar{\pi}_{3456}^{3i}$. This shows that the original points whose representation is in Fig. 5.47 (left) are on a 5-flat in \mathbb{R}^6 . The remaining points of the representation are obtained in the same way, and all seven points of the representation of π^5 are seen on the right. The coefficients of its equation are equal to six times the distance between sequentially indexed points as in Fig. 5.18 for \mathbb{R}^3 .

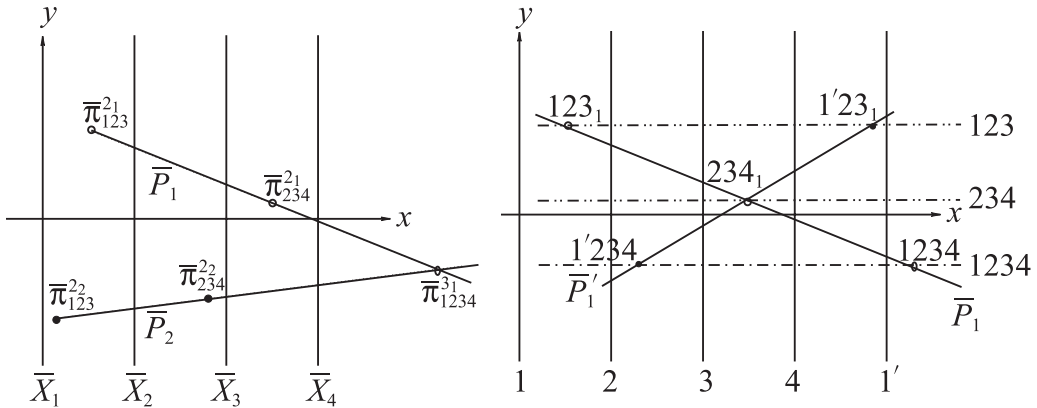
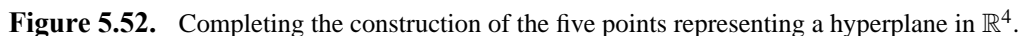
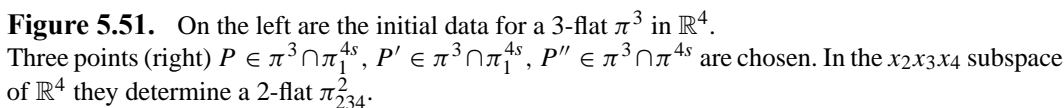


Figure 5.50. Continuing the construction of indexed points for a 3-flat π^3 in 4D.

On the left is the point $\bar{\pi}_{1234}$ previously constructed in Fig. 5.46. Using simplified notation, the construction is continued on the right to obtain the point $\bar{\pi}_{1'234}$, marked by $1'234$. The points $123_1, 1'23_1, 234_1$ are points of the representation of a 2-flat π_1^2 contained in π^3 . The lines \bar{P}_1, \bar{P}'_1 on 1234 and $1'234$ share the indices 234 and necessarily intersect at the point 234_1 .



The coplanarity of a set of points $S \subset \pi$ can be visually verified. If the points are perturbed, staying close to but no longer being on the plane π , can “near coplanarity” still be detected? Let us formulate this question more specifically by perturbing the coefficients c_i of a plane’s equation by a small amount ϵ_i . This generates a family of “proximate” planes forming a surface resembling a “multiply twisted slab.” Now the experiment is performed by selecting a random set of points from such a twisted

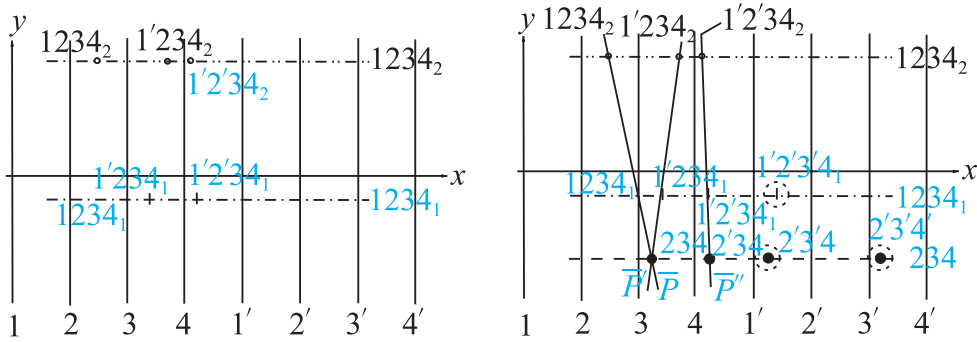


Figure 5.53. On the left are the initial data specifying two 3-flats π^{3_1}, π^{3_2} in \mathbb{R}^4 . On the right, the lines $\bar{P}, \bar{P}', \bar{P}''$ are drawn on the pair of points 1234, 1'234, 1'2'34 respectively, providing the 234 and 2'34 points for the 2-flat $\pi^2 = \pi^{3_1} \cap \pi^{3_2}$. Then the points 2'3'4' and 1'2'3'4', shown within the dotted circles, are constructed.

slab, and repeating the construction for the representation of planes. As shown in Figs. 5.55, 5.56, there is a remarkable resemblance to the coplanarity pattern. The construction also works for any N . It is also possible to obtain error bounds measuring the “near coplanarity” [131]. This topic is covered in Chapter 8 on proximity.

Experiments performing similar constructions with points selected from several twisted slabs simultaneously showed that it is possible to determine the slabs from which the points were obtained or conversely can be fitted to. All this has important and interesting applications (USA patent # 5,631,982). Given a set of points composed of point clusters each from a different plane (or hyperplane), the determination of these planes can be made with very high accuracy and low computational complexity, as shown in the appendix on *recent results*.

5.7 Representation Mapping, Version II

Let us revisit and update the representation mapping

$$\mathcal{J} : 2^{P^N} \rightarrow 2^{P^2} \times 2^{[1,2,\dots,N]}, \quad (5.92)$$

in view of this chapter’s results. The recursive construction starts with the *non-recursive step* for the representation of a point $P \in \mathbb{R}^N$, a 0-flat π^0 , in terms of N points each with one index, the values of the coordinates $p_i, i \in [1, \dots, N]$, of P on the \bar{X}_i axes. For consistency, a point is also specified by the N equations

$$x_i = p_i, \quad i = 1, \dots, N. \quad (5.93)$$

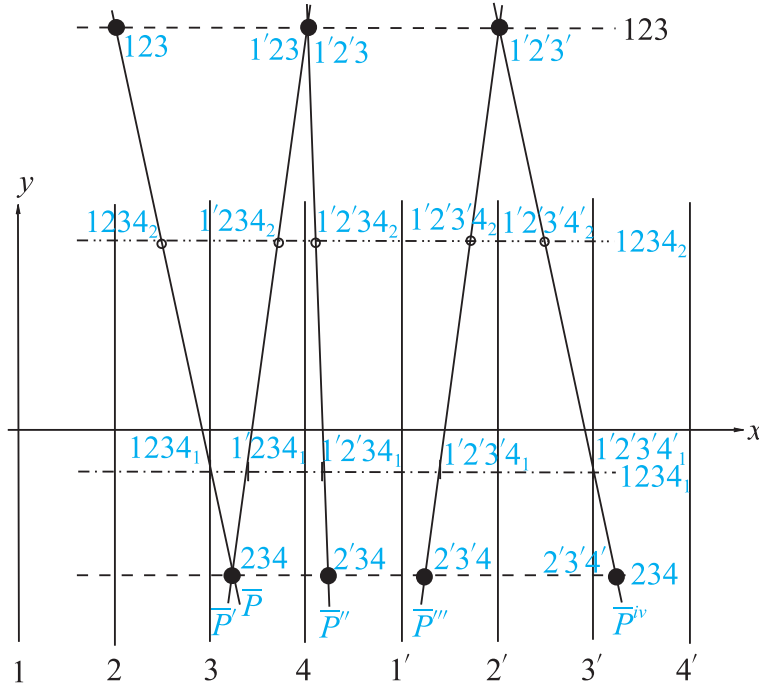


Figure 5.54. Intersection of two 3-flats in 4-D.

The result is the 2-flat π^2 given by the points 123 , $1'2'3'$, etc. representing π_{123}^2 , and 234 , $2'34$, etc. for π_{234}^2 .

The construction of a p -flat from its points (0-flats) to lines (1-flats) and so on proceeds according to Theorem 5.4.4, the dimensionality being raised by 1 until it reaches p , providing the indexed points needed for the representation.

Let N_r be the number of points and n_r the number of indices appearing in the representation of a flat in \mathbb{R}^N . Then for a p -flat, $N_r = (N - p)p$, $n_r = p + 1$, so that

$$N_r + n_r = (N - p)p + (p + 1) = -p^2 + p(N + 1) + 1. \quad (5.94)$$

In \mathbb{R}^N some examples are

1. $p = 0$: points π^0 : $N_r + n_r = N + 1$,
2. $p = 1$: lines π^1 : $N_r + n_r = (N - 1) + 2 = N + 1$,
3. $p = 2$: 2-flats (2-planes) π^2 : $N_r + n_r = (N - 2)2 + 3 = 2N - 1$,
4. $p = N - 1$: hyperplanes $N_r = N - 1$, $n_r = N$: $N_r + n_r = N - 1 + N = 2N - 1$.

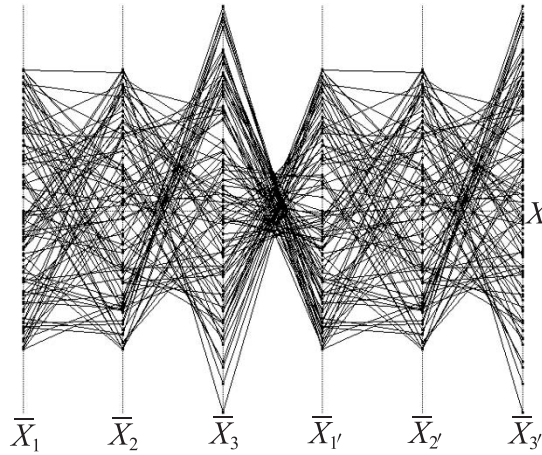


Figure 5.55. Polygonal lines representing a randomly selected set of “nearly” coplanar points.

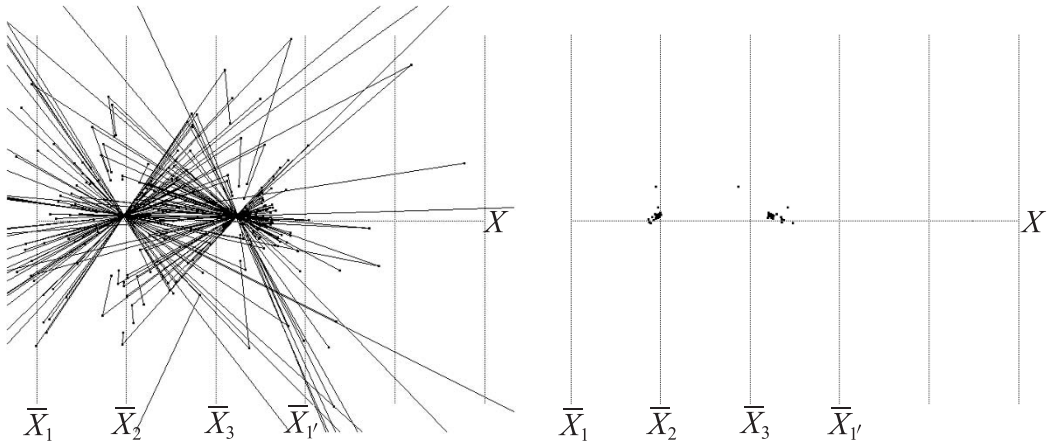


Figure 5.56. The “near-coplanarity” pattern.

The picture on the left is very similar to that obtained for coplanarity, with the points of intersection forming two clusters (right).

Note that (5.94) does not include the case for $p = 0$ (points). In summary, the representation by points reveals that the object being represented is a p -flat whose dimensionality is one less than the number of indices used. The dimensionality of the space where the p -flat resides is, of course, equal to the number of parallel axes, or it can also be found from the number of points $(N - p)p$ if one knows the value of p .

We have seen that the image $\mathcal{I}(U) = \bar{U}$ of a $U \subset \mathbb{P}^N$ consists of several points each *indexed* by a subset of $[1, 2, \dots, N]$, the indexing being an essential part of the representation. The mapping \mathcal{I} provides a unique representation, since it is one-to-one with $\bar{U}_1 = \bar{U}_2 \Leftrightarrow U_1 = U_2$. By representation is meant a *minimal* subset of points that uniquely identify the object in question, for as we have seen, there are many redundant points. The final version of the \mathcal{J} is given in the conclusion after the representation of surfaces.

** Envelopes

Note: Reading of this chapter is discretionary. It provides foundational background for the chapters on curves and surfaces.

6.1 The Basic Idea

Envelopes are encountered in many areas of mathematics as well as in several applications. Also, they are useful background for the construction of representations in parallel coordinates. In our era, envelopes are no longer a “fashionable” subject and are often not covered in coursework. For good references one needs to look at older sources. For these reasons, a quick overview is included here. An extensive treatise is available in [17], while good shorter sections can be found in [35], [73], and [143].

First let us build some intuition. Consider a line $\ell(x, y, a) = 0$ in the xy plane, as shown in Fig. 6.1, rotated at a constant distance r from a point O . Its position is specified by the angle a . This rotation traces a circle C of radius r and centered at O . At each point of C there is exactly one line, as specified by a , that is tangent to C at that point. We say that C is the

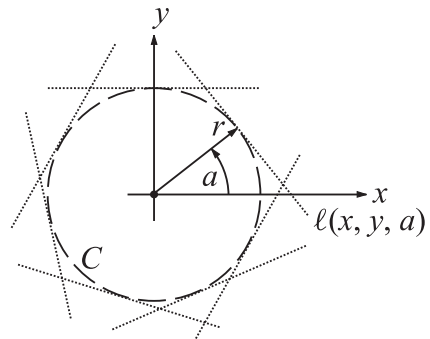


Figure 6.1. Envelope is a circle.

The circle C is tangent at each point to a specific line $\ell(x, y, a)$, which is rotated at a constant distance r from the point O .

envelope of the family of curves $\ell(x, y, a) = 0$. For another example, consider a rod with length d constrained in the vertical plane xy sliding so that it always touches the wall and floor as shown in Fig. 6.2. Formally, we have a family of segments $d(x, y, a)$ all having constant length d whose position can be prescribed by the angle a between the floor and the segment. These rods are tangent to a curve called A , which is one of the four arcs of an *astroid* given by

$$x^{2/3} + y^{2/3} = d.$$

This, of course, needs to be shown.

In each example there is a one-parameter family, \mathcal{F} , of curves $F(x, y, a) = 0$, each curve being specified by a value of the parameter a . Also, in each case we saw that there is yet another curve \mathcal{E} that is tangent at each point to a curve in the family \mathcal{F} . In the first case this was the circle C , and in the second, the astroid A . It happened that the families of curves in both cases were linear in x and y , but this does not detract from the generality. For example, the astroid, with all four arcs, turns out to be tangent at each one of its points to a curve of the family

$$F(x, y, a) = \frac{x^2}{a^2} + \frac{y^2}{(1-a)^2} - 1 = 0,$$

as shown in Fig. 6.3. These are the ellipses centered at the origin with semiaxes of length a and $a - 1$. That is, the sum of the length of the semiaxes is always 1; so the circle with diameter 1 is included in this family.

6.2 Formulation

A one-parameter family of curves in the xy plane can be described by a function

$$F(x, y, a) = 0, \tag{6.1}$$

where $a \in R$. There may exist a curve \mathcal{E} that touches²⁸ at each of its points a member of (6.1). The curve \mathcal{E} is called the *envelope* of the family \mathcal{F} . There remains to see

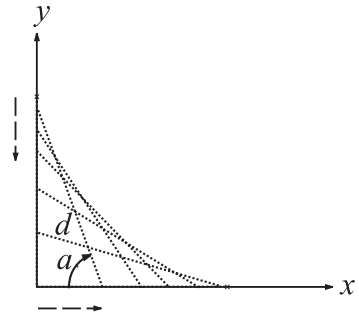


Figure 6.2. Envelope is one quarter of an *astroid*.

A rod d constrained to slide in the vertical plane while touching both the vertical wall and the floor traces (i.e., is always tangent to) a curve called an *astroid*.

²⁸In the ensuing it will become clear why the word “touches” rather than “tangent” is used here.

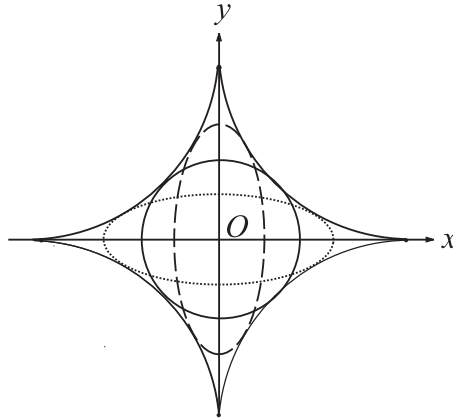


Figure 6.3. Envelope is a full *astroid*.

The envelope of the family of ellipses with sum of their semiaxes constant.

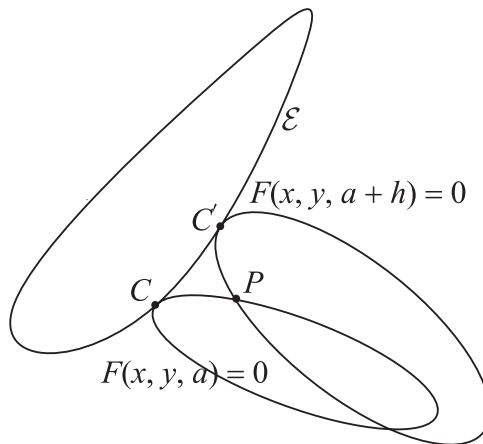


Figure 6.4. The envelope \mathcal{E} and two intersecting neighboring curves of F .

how \mathcal{E} , when it exists, can be found. Keeping with the mantra “let intuition guide the formalism,” we will add the conditions as the need is encountered and explained. At first we assume that the curves of \mathcal{F} are *tangent* to \mathcal{E} , and that neighboring curves of the family intersect at points “close” to \mathcal{E} and lie on the same side. This is the purpose of the illustration in Fig. 6.4, where the curves $F(x, y, a) = 0$ and $F(x, y, a + h) = 0$ are shown intersecting at P . The idea is to find the point of intersection P of these two curves and let $h \rightarrow 0$. Then, given sufficient smoothness

conditions, P must approach the point of contact of $F(x, y, a) = 0$ with \mathcal{E} , since at P ,

$$F(x, y, a) = F(x, y, a + h) = 0,$$

and therefore

$$\frac{F(x, y, a + h) - F(x, y, a)}{h} = 0.$$

When the function $F(x, y, a) = 0$ is differentiable with respect to $a \forall a \in R$ and $h \rightarrow 0$ above, we obtain that the point of contact with \mathcal{E} has coordinates (x, y) , which are found by eliminating the parameter a from the two equations:

$$\begin{cases} F(x, y, a) = 0, \\ \frac{\partial F}{\partial a}(x, y, a) = 0. \end{cases} \quad (6.2)$$

6.3 Necessary and Sufficient Conditions

Under what conditions does (6.2) yield the envelope, and what other possibilities may exist? Let us assume that \mathcal{E} is parametrizable and can be given in terms of two continuously differentiable functions

$$x = x(a), \quad y = y(a), \quad (6.3)$$

with

$$\left(\frac{dx}{da}\right)^2 + \left(\frac{dy}{da}\right)^2 \neq 0, \quad (6.4)$$

and is tangent to the curve $F(x, y, a) = 0$ with the *same* value of the parameter a at the point of contact, as shown in Fig. 6.5. The underlying reason why such a parametrization is possible here is that at each point (x, y) of \mathcal{E} there exists an a specifying $F(x, y, a)$ at that point. Hence it is possible to parametrize \mathcal{E} in terms of a with

$$F(x(a), y(a), a) = 0 \quad (6.5)$$

valid at the point of contact. Differentiating (6.5) with respect to a , we obtain

$$F_x \frac{dx}{da} + F_y \frac{dy}{da} + F_a = 0. \quad (6.6)$$

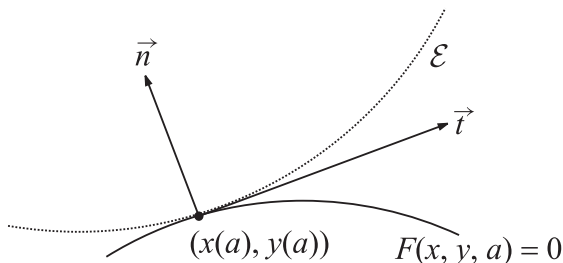


Figure 6.5. Point of contact between the envelope \mathcal{E} and a curve in F .

The derivatives $\frac{dx}{da}, \frac{dy}{da}$ are proportional to the direction cosines of the tangent \vec{t} of \mathcal{E} , while the quantities F_x, F_y are proportional to the direction cosines of the normal \vec{n} to the curve $F(x, y, a) = 0$ of the family. These two vectors \vec{t}, \vec{n} must be orthogonal to each other. Therefore, the inner product of the corresponding direction cosines satisfies

$$F_x \frac{dx}{da} + F_y \frac{dy}{da} = 0, \quad (6.7)$$

so that (6.6) and (6.7) together imply

$$F_a(x, y, a) = 0.$$

Hence the conditions of (6.2), which were previously found *sufficient*, have now been shown to be *necessary* for the existence of \mathcal{E} .

6.3.1 Singular Points

When

$$F_x(x, y, a) = F_y(x, y, a) = 0,$$

the curve $F(x, y, a) = 0$ of the family may have *singular points* (where F is not continuously differentiable there), and no conclusions can be drawn about the contact of the curves. For algebraic curves (i.e., polynomials), the singular points may be

- points of *self-intersection*: points where the curve crosses itself,
- points of *self-tangency*: where the derivative is continuous but the curve has portions above and below the tangent at each point,
- *cusps*: where the derivative is not continuous,
- *isolated points*: points disjoint from the curve.

An algebraic curve can have only a finite number of singular points, and when it has none, it is called *smooth*. When the envelope is found using (6.2), \mathcal{E} may also contain singular points.

6.4 Examples: Envelopes of Families of Curves

The following examples are intended to clarify the method and indicate various possibilities.

1. The family

$$F(x, y, a) = x^2 + y^2 - 2x \cos a - 2y \sin a = 0 \quad (6.8)$$

describes a family of circles that have unit radius and pass through the origin. This becomes a bit clearer when (6.8) is rewritten as

$$(x - \cos a)^2 + (y - \sin a)^2 = 1,$$

indicating that the circles are centered on the circumference of the circle

$$C_1 : x = \cos a, y = \sin a,$$

which has unit radius and center at the origin, as shown in Fig. 6.6.

To find the envelope, we differentiate to obtain

$$F_a = 0 \implies x \sin a - y \cos a = 0. \quad (6.9)$$

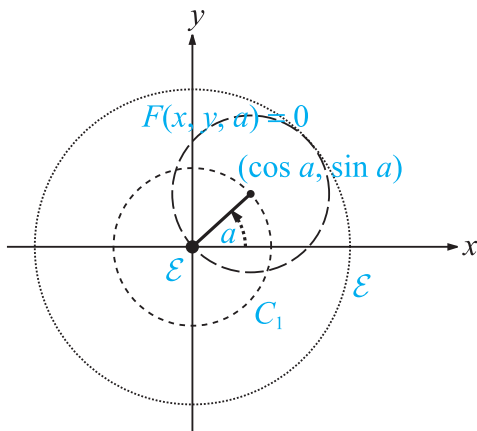


Figure 6.6. Family of circles centered on the unit circle C_1 . Note the envelope consisting of the outer circle and the isolated point at the origin.

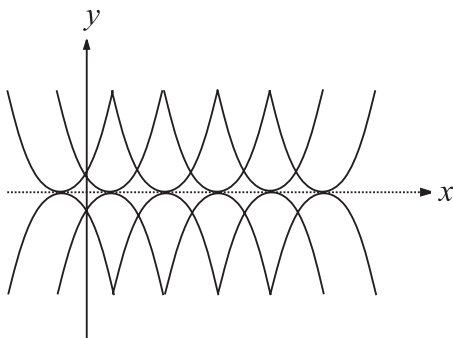


Figure 6.7. Family of self-tangent curves and their envelope.

The two equations (6.8) and (6.9) are satisfied by $x = 0$ and $y = 0$. Hence $(0, 0) \in \mathcal{E}$. Further, when $x^2 + y^2 \neq 0$, the solution (do this as an exercise) is

$$\sin a = \frac{y}{2}, \quad \cos a = \frac{x}{2},$$

and eliminating a yields

$$x^2 + y^2 = 4.$$

That is, using (6.2), we found that \mathcal{E} consists of

- (a) the circle of radius 2 centered at the origin, and
- (b) the *isolated point* at the origin.

2. Now we consider the family of curves

$$F(x, y, a) = y^2 - (x - a)^4 = 0. \quad (6.10)$$

Each curve in this family consists of a parabola and its reflection about the x axis, so there is a point of *self-tangency* at $(a, 0)$. Using (6.2), the envelope is found to consist of all the points (called the *locus*) of self-tangency, as illustrated in Fig. 6.7. This example shows that it is not essential, as assumed in the formulation, for the curves to be on *one side of* \mathcal{E} , and it suffices that there be sets of points of intersection of neighboring curves on either side that enable the limit as $h \rightarrow 0$ in the derivation of (6.2) to exist.

3. Applying (6.2) to the family of cubical parabolas

$$F(x, y, a) = (x - a)^2 - y^3 = 0 \quad (6.11)$$

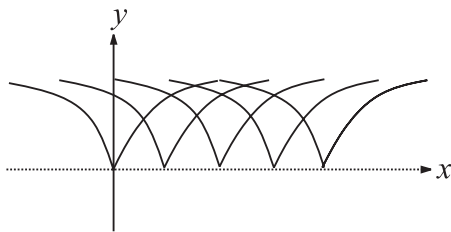


Figure 6.8. Family of curves with cusps.

gives the x axis, which is the *locus of the cusps*, as a result. From Fig. 6.8 it is clear, however, that this is not the envelope, since it is not *tangent* to the curves of the family at the point of contact. It is instructive to review the derivation of (6.2) and find which of the assumptions failed to be true here.

4. The next example illustrates that the points of the envelope need not be the limit of neighboring intersecting points of the curves in the family. Consider the family of cubical parabolas

$$F(x, y, a) = y - (x - a)^3 = 0, \quad (6.12)$$

which are identical except for the translation by a and do not intersect each other. Applying (6.2) provides the condition $F_a = 3(x - a)^2 = 0$, which in turn shows that \mathcal{E} is the x axis. We accept this as the envelope of the family, since each F is tangent to it, as shown in Fig. 6.9.

5. For our last example we consider the family of *strophoids*

$$F(x, y, a) = [(y - a)^2 + x^2](x - 2) + x = 0, \quad (6.13)$$

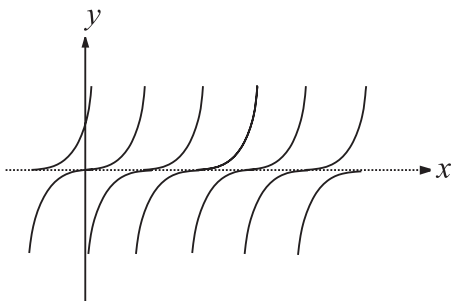


Figure 6.9. Family of curves with inflection points.

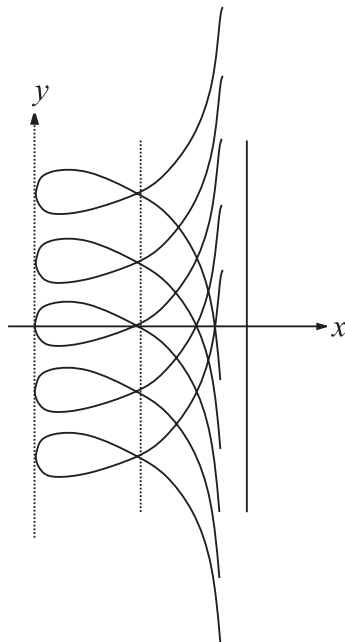


Figure 6.10. Family of self-intersecting curves.

which as Fig. 6.10 shows are self-intersecting (the point of intersection is also called a *double point*). Each curve can be translated to any other curve. We find that

$$F_a = -2(y - a)(x - 2) = 0 \implies x = 2 \text{ or } y = a.$$

Since no finite value of y corresponds to $x = 2$, we take $y = a$, so that $x^2(x - 2) + 2 = 0$. This suggests that \mathcal{E} consists of the two vertical lines $x = 0$ and $x = 1$. Yet it is clear from the figure that only $x = 0$ can be considered as the envelope, while the line $x = 1$ is the locus of the self-intersections. Again it is a good idea to run through the derivation to see which assumptions failed to be true.

Exercises

1. Prove that one arc of the *astroid* is the envelope of the falling stick as described at the beginning of this section.
2. Prove that all four arcs of the astroid are the envelopes of the ellipses whose semi-axes have a constant sum.

3. Find the envelope of the family of ellipses whose semiaxes are collinear and whose areas are constant. Hint: Area of ellipse = πab , where a, b are the lengths of the semiaxes.

In connection with parallel coordinates,

4. Prove the basic point \leftrightarrow line duality using envelopes.
5. Derive the vertical-line representation of planes using envelopes.
6. Derive the 2-point representation of planes in R^3 using envelopes.
7. Show using envelopes that in P^2 , conics are mapped into conics. This is an excellent but challenging project topic.

7

Curves

7.1 **FT-1** Point-Curves and Line-Curves

We revisit the fundamental duality in the plane, Chapter 3, starting with the point-to-line correspondence

$$P : (p_1, p_2, p_3) \longrightarrow \bar{P} : [(p_1 - p_2), dp_3, -dp_1], \quad (7.1)$$

where d is the distance between the \bar{X}_1 and \bar{X}_2 axes and the triples within $[\cdot \cdot \cdot]$ and $(\cdot \cdot \cdot)$ are the line and point homogeneous coordinates respectively. Unless otherwise stated, P is a regular (i.e., Euclidean) point, so without loss of generality, $p_3 = 1$, whence

$$P : (p_1, p_2, 1) \longrightarrow \bar{P} : [(p_1 - p_2), d, -dp_1]. \quad (7.2)$$

The line-to-point correspondence is

$$\ell : [a_1, a_2, a_3] \longrightarrow \bar{\ell} : (da_2, -a_3, a_1 + a_2), \quad (7.3)$$

for $\ell : a_1x_1 + a_2x_2 + a_3 = 0$. When $a_2 \neq 0$, the slope of ℓ is $m = -\frac{a_1}{a_2}$, the intercept is $b = -\frac{a_3}{a_2}$, and

$$\ell : [m, -1, b] \longrightarrow \bar{\ell} : (d, b, 1 - m). \quad (7.4)$$

The image of each point P on a curve c is a line obtained from (7.1), as shown in Fig. 7.1 (left), and the *envelope* of all such lines, when it exists, is the curve \bar{c} , the image of c . So we need to consider two types of curves: the *point-curves* we have known and loved all our lives, and which are made up of points, and *line-curves*,

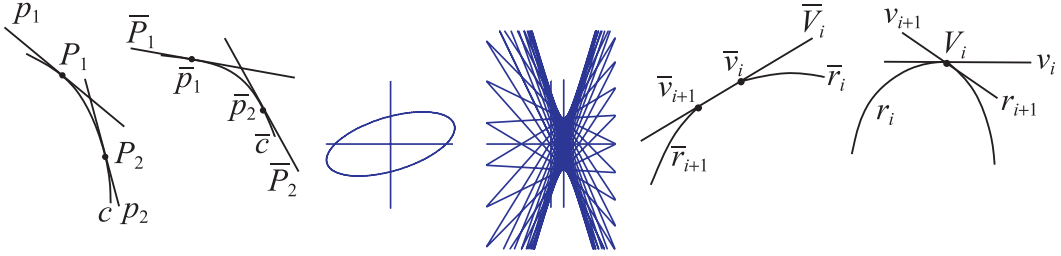


Figure 7.1. Point c and line curves \bar{c} . (Right) Vertex in r maps into a straight line in \bar{r} .

which are formed by the envelope of a family of lines, usually but not always their tangents.

The *point* \leftrightarrow *line* duality in the projective plane \mathbb{P}^2 causes the duality

$$\text{point-curve} \leftrightarrow \text{line-curve}.$$

In fact, $\bar{\ell}$, the point image of the line (curve) of ℓ given by (7.3), is the *envelope* of the lines \bar{P} for $P \in \ell$. In this chapter we learn to construct the representation of curves in $\|\text{-coords}$ efficiently and study properties of the image curves. In the ensuing, we consider piecewise differentiable curves r , i.e., $r = r_1 \cup r_2 \cup \dots \cup r_i \cup r_{i+1} \dots \cup r_n$, where the r_i are differentiable. The endpoints of r may be ideal. At a vertex $V_i = r_i \cap r_{i+1}$ where there are two tangents, \bar{r} will contain a straight-line segment \bar{V}_i separating the two line-curves \bar{r}_i, \bar{r}_{i+1} ; see Fig. 7.1 (right).

7.1.1 Separation in the xy Plane

Recall the basic idea of separating points by a line. In the Euclidean x_1x_2 plane, a point $P : (p_1, p_2)$ is said to be on, above, or below the line $\ell : x_2 = mx_1 + b$ if the expression $(p_2 - p_1m) =, <, > b$; see the right-hand part of Fig. 7.2. This is equivalent to placing a normal vector \vec{n} on the line ℓ , providing two half-planes on either side of the line as well as a direction (that of the normal). The above or below relation of

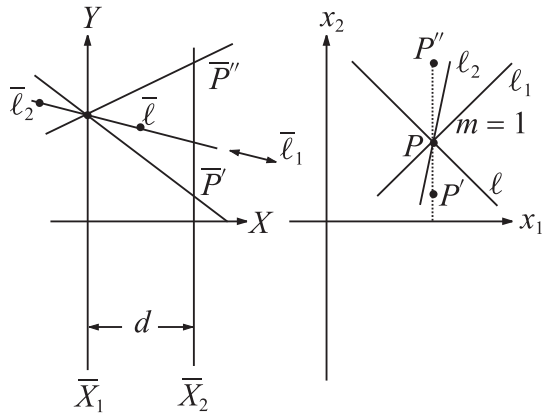


Figure 7.2. The “above” and “below” relations between points and lines switch at $m = 1$.

a point P with respect to ℓ is then specified in terms of the half-plane on which P lies. For a vertical line, say $x_1 = c$, the direction still determines unambiguously the half-spaces and hence the above/below relation, which would otherwise be ambiguous.

Alternatively, we can revert to the line (7.3) with $a_1 = 0$ and the convention that the direction of increasing x_1 specifies the “above” half-plane. Since we are working in the projective plane \mathbb{P}^2 , we consider any regular (i.e., Euclidean) point as being below the ideal line ℓ_∞ . As indicated on the left-hand side of Fig. 7.2, this situation is more fun in $\|\text{-coords}$, for there the separation criterion “flips” at $m = 1$. Correspondingly, \bar{P} is the line $y = (p_2 - p_1)x + p_1$, where to simplify matters we set $d = 1$ from now on. The various cases for $\ell \neq \ell_\infty$ are summarized²⁹ in the following lemma.

- Lemma 7.1.1.** 1. P is on, below (above) a line ℓ whose slope $m < 1 \iff \bar{P}$ is on, below (above) $\bar{\ell}$.
2. P is on, below (above) a line ℓ whose slope $m \geq 1 \iff \bar{P}$ is on, above (below) $\bar{\ell}$.

Let

$$M_c(I) = \max\{m_c(P) \mid P(x_1) \in c, x_1 \in I\}, \quad (7.5)$$

where $m_c(P)$ is the slope of the tangent of the curve $c : x_2 = r(x_1)$ at the point P . Further, by considering at each vertex the supporting line with maximum slope instead of the tangent in the above definition, $M_c(I)$ can be defined for a complete piecewise smooth curve. Clearly, then, tracing the above/below relation in the image \bar{c} of a curve c depends on whether $M_c(I)$ goes through the value 1. There are some straight forward consequences of the lemma, which for convenience are listed separately. They are true because the statements are true pointwise.

- Corollary 7.1.2.** 1. For $M_c < 1$, c is below (above) a point $P \iff \bar{c}$ is below (above) \bar{P} .
2. For $M_c \geq 1$, c is below (above) a point $P \iff \bar{c}$ is above (below) \bar{P} .

Definitio 7.1.3. A curve segment r is convex downward (cd) (also sometimes called “concave”) if it is on or below any one of its tangents (or more generally, supporting lines). It is convex upward (cu) if it is on or above any one of its tangents; see Fig. 7.3.

When for the curve $r : x_2 = r(x_1)$, r' and r'' exist in the interval of interest, then where r is cd, r' is monotone increasing; hence $r'' > 0$. Further, where r is cu,

²⁹Fortunately, consideration of the “slope subcases” is not required for higher dimensions; see section 5.3.5 in chapter 5 on planes.

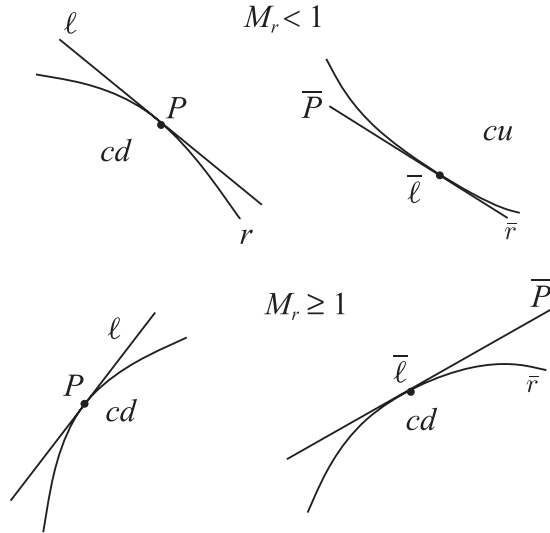


Figure 7.3. Curves r flip $cu \leftrightarrow cd$ for $M_r < 1$ and do not flip for $M_r \geq 1$; Corollary 7.1.4.

r' is monotone decreasing, so $r'' < 0$. If there is a transition from cu to cd (or vice versa), then necessarily $r'' = 0$ there; more on this shortly.

Corollary 7.1.4. 1. If r is a cd (cu) curve segment with $M_r < 1$, then \bar{r} is cu (cd).

2. If r is a cd (cu) curve segment with $M_r \geq 1$, then \bar{r} is cd (cu).

Proof. Since \bar{r} is on or above any of its tangents, it follows that \bar{r} cu , as shown in Fig. 7.3. Pay attention that even though the above/below relation between ℓ , P , $\bar{\ell}$, and \bar{P} is preserved, the cu/cd property between r and \bar{r} “flips,” since r is a point-curve and \bar{r} is a line-curve. The proof for the remaining cases (two of which are illustrated in Fig. 7.3) is completely analogous. ■

Certain properties are not intrinsic to a curve. Rather, they are incidental to the curve’s position or orientation with respect to a specific coordinate system. For example, vertical or horizontal tangents do not persist under a rotation, and for that matter, $cu \leftrightarrow cd$ with appropriate rotations. Properties such as shape (circle, triangle) that are unaltered by rotations, translations, and scaling are geometrical. The properties of curves and surfaces varying pointwise are studied in *differential geometry* using the techniques of calculus and vector analysis for the more elementary aspects, and tensor methods for the more advanced. In keeping with our aim of making this text widely accessible, only the essentials of this subject are employed. For a more rigorous and thorough background there are excellent references such

as [166], [75], [183], [48], with [121] recommended for a most enjoyable overview interlaced with the historical development of differential geometry (Chapters 23 and 37).

7.2 Duality between Cusps and Inflection Points

The description of *inflection points* (ip) turns out to be a bit stickier than that given at the elementary calculus level (see, for example, [55] and [180]). The *tangent* to a curve c touches it at the point of contact $P = (x_1, r(x_1))$. Where $r''(x_1) \neq 0$, the circle with the largest radius $R(x_1)$ having the same tangent at P as c is found, and it measures the local *curvature* $= 1/R$ of c (though “circularity” would be a more accurate term). The circle’s arc is the best “circular” approximation of the curve at P . Clearly this circle’s center, called the center of curvature, lies on the

concave side of c . The directed line perpendicular to the tangent at P and pointing toward the center of curvature is the *normal*. Directing the tangent so that together with the normal as axes they form a right-handed orthogonal coordinate system provides a convenient frame of reference for the study of the curve’s behavior and characterization of various types of points, described next and illustrated in Fig. 7.4.

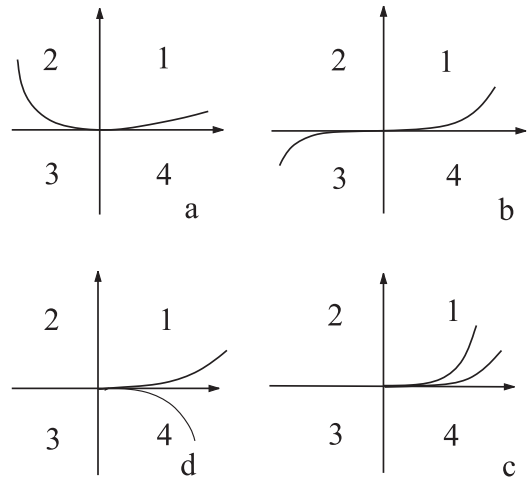


Figure 7.4. Types of points on a curve.

The coordinate system here is the tangent and normal at the point.

1. *Regular*³⁰: where the normal but not the tangent crosses the curve. A line crosses a curve when parts of the curve (consisting of more than a point) lie on both sides of the curve as in (a) of Fig. 7.4.
2. *Inflection point* (ip): where both the tangent and the normal cross the curve as in (b) of Fig. 7.4.
3. *Cusp* of the first kind: where the tangent but not the normal crosses the curve as in (d) of Fig. 7.4.

³⁰This name here is not to be confused with a regular (i.e., Euclidean) point of the projective plane.

4. *Cusp* of the second kind: where neither the tangent nor the normal crosses the curve as in (c) of Fig. 7.4.

This by no means exhausts the possibilities. We will not deal with other kinds of singularities, which have more complicated rigorous descriptions, though some are easy enough to illustrate. Points that are not regular are called *singular*. In fact, we will consider only cusps of the first kind and refer to them simply as cusps. It is important to note that curves having cusps and inflection points have both a cu and cd portion in a neighborhood of these points. Strikingly, our point–line duality induces an ip–cusp duality. This result, which we obtain next with straight forward geometric argument, is more general than that shown in algebraic geometry specifically for algebraic (polynomial) curves [61, p. 88].

Theorem 7.2.1 (ip \leftrightarrow cusp duality) ♣ FT-2

1. *I* ip of a point-curve r with tangent $i \leftrightarrow \bar{i}$ cusp of the line-curve \bar{r} with tangent \bar{I} .
2. *C* cusp of a point-curve r with tangent $c \leftrightarrow \bar{c}$ ip of the line-curve \bar{r} with tangent \bar{C} .

Proof. Step 1. Consider a curve r having an inflection point I with tangent i in the tangent (x_1)–normal (x_2) coordinate system shown in Fig. 7.5. Since the slope of the tangent i at I is zero, \exists an interval S on the x_1 (tangent) axis containing I and where $M_r < 1$. It is easy to see that in a neighborhood S of an inflection point the curve is necessarily monotone (increasing or decreasing). Further, with S , r has a

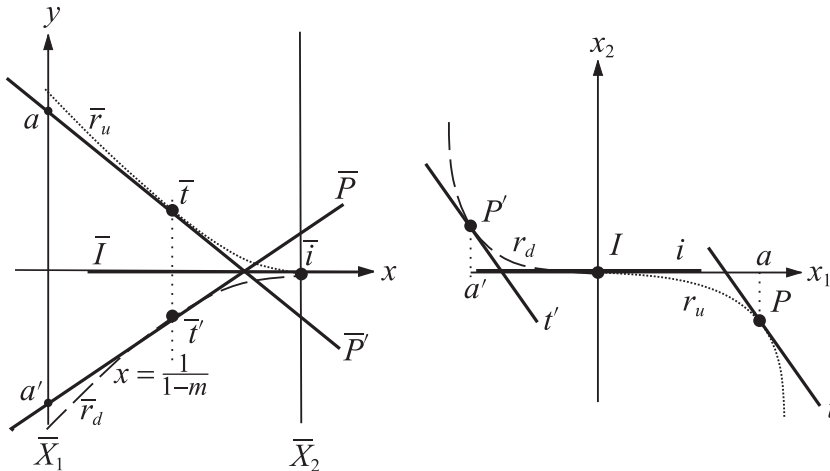


Figure 7.5. Inflection point (ip) to cusp.

Step 5. Now let us consider the case that the tangent at the point of inflection has slope other than 0. Specifically, as shown in Fig. 7.6, r has inflection point I_1 with tangent i_1 having slope less than 1 and not necessarily 0. By exactly the same argument as above, \bar{r} has a cusp at \bar{i}_1 with tangent \bar{i}_1 . The portions r_u and r_d become $cd \bar{r}_u$ and $cu \bar{r}_d$ respectively.

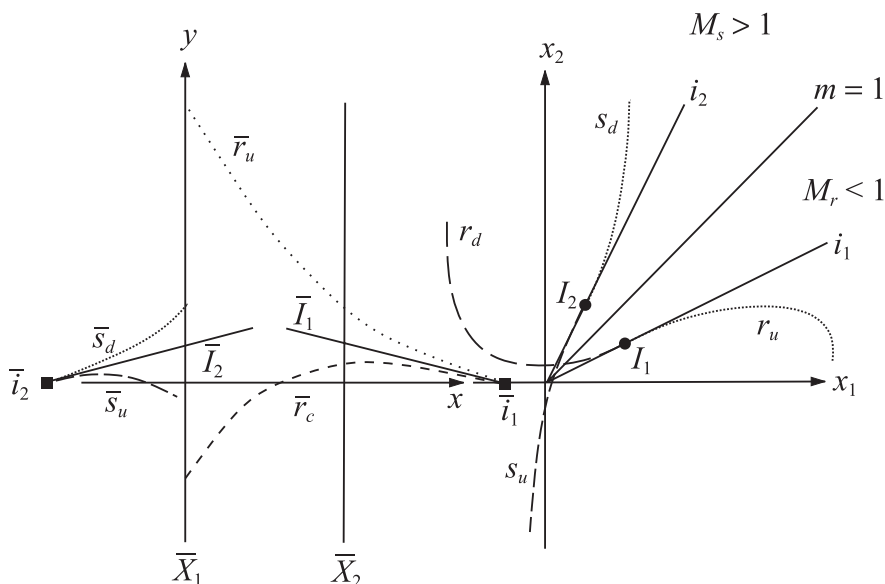


Figure 7.6. The duality inflection points \leftrightarrow cusps is slope-independent $m \neq 1$.

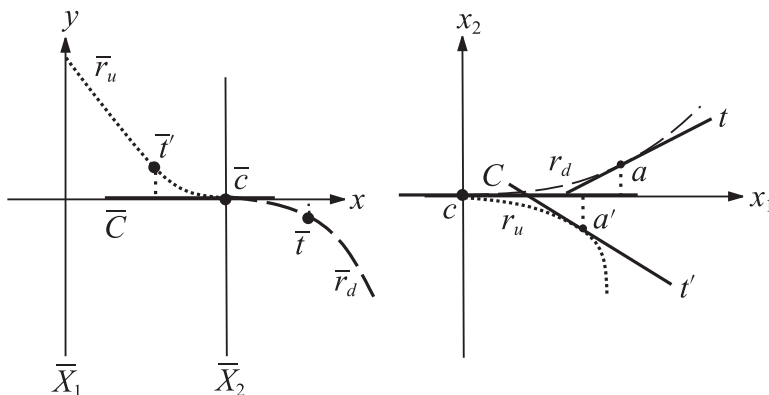


Figure 7.7. The image of a cusp is an inflection point.

Step 6. The counterpart for a curve s with $M_s > 1$ having an inflection point (see Fig. 7.6) is still the same kind of argument, where now the cu and cd parts stay cu and cd in the image \bar{r} .

Step 7. The proof of the second assertion for cusps is similar. Again we start with a point-curve r shown in Fig. 7.7 with the normal–tangent coordinate system, having a cusp C with slope c . It is clear that the image \bar{c} equals $x \cap \bar{X}_2$. Since C is a cusp, \exists an interval S on the x_1 (tangent) axis with $C \in S$, and on this interval, $M_r < 1$. Further, $\forall a \in S \exists a' \in S$ such that the derivatives $r'(a)$, $-r'(a')$ are equal, with t and t' being the tangents to r at a and a' respectively. This is due to the double-valuedness of r in some neighborhood.

Step 8. Step 7 implies that \bar{t} and \bar{t}' will be on opposite sides of the \bar{X}_2 axis where \bar{c} resides. Hence r_u and r_d will be on opposite sides of the \bar{X}_2 axis, since this is true $\forall a \in S$. Since $M_r < 1$ there, \bar{r}_u is cd and \bar{r}_d is cu.

Step 9. The image $\bar{r} = \bar{r}_u \cup \bar{r}_d$ by Step 8 has an inflection point at \bar{c} with tangent \bar{C} .

Step 10. The proof when the tangent at the cusp is not horizontal is the exact counterpart of the proof for the inflection given above and illustrated in Fig. 7.8.

Step 11. Finally a word about the slope of the curve going through 1. The case for inflection points is illustrated in Fig. 7.9. The case for cusps involves a subtlety, and the reader is encouraged to work it out in detail (see exercises below). ■

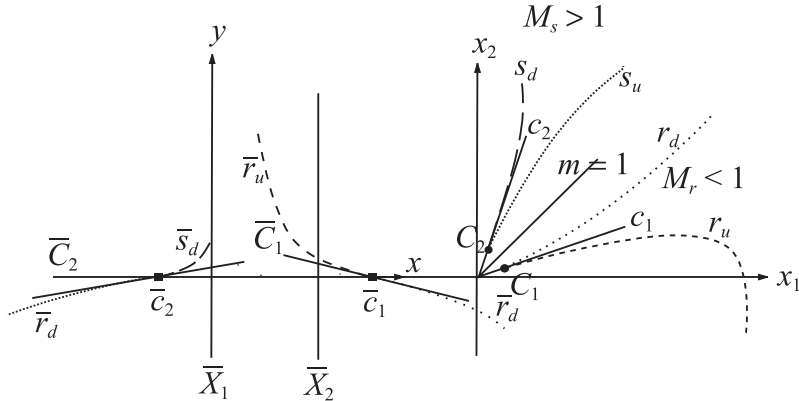


Figure 7.8. The duality cusp \leftrightarrow inflection point is independent of slope.

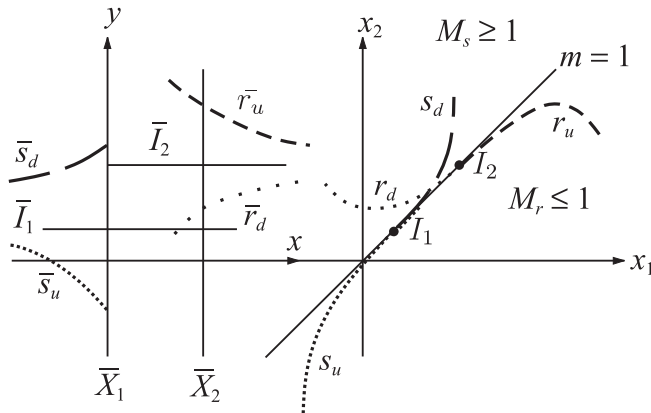


Figure 7.9. Inflection points to cusps when the curve goes through $m = 1$.

Exercises

1. Show that in a neighborhood of an inflection point I of a curve r :
 - (a) the curve r has a cu portion r_u and cd portion r_c separated by I , and
 - (b) the curve r is monotone there.
2. Show that in a neighborhood of a cusp C of a curve r :
 - (a) the curve r has a cu and cd portion separated by the tangent at C , and
 - (b) the curve r has a monotone increasing and monotone decreasing portion also separated by the tangent at C .

3. Find the envelope of each of the following curves in a neighborhood of its singular points (most of them at $(0, 0)$) and identify the kind of singularity:
 - (a) $y = x^3$.
 - (b) $(x^2 + y^2)^2 - 2a^2(x^2 - y^2) = 0$.
 - (c) the *astroid*: $x^{2/3} + y^{2/3} = a^{2/3}$.
 - (d) $y = x^{1/3}$.
 - (e) $y^3 - x^2 = 0$.
 - (f) the *cisoid*: $x^3 + xy^2 - 2ay^2 = 0$.
 - (g) $y = x^3, x < 0$ and $= x^2, x \geq 0$.
 - (h) the *arclight curve*: $x^2y^2 + a^2x^2 - a^2y^2 = 0$.
 - (i) $y = x^{2/3}$.
 - (j) $y^3 - x^4 = 0$.
 - (k) the *trisectrix*: $x^3 + xy^2 + ay^2 - 3ax^2 = 0$.
4. Carefully study the image of cusps whose tangent has slope 1.
5. What happens when a point-curve has vertical slope at its cusp, and what is the point-curve whose line-curve image has a cusp with vertical slope?

7.3 FT-3 Point-Curves \rightarrow Point-Curves

As posed, the construction of a curve's image involves the sequence of operations

$$\{\text{point-curve}\} \rightarrow \{\text{line-curve}\} \rightarrow \{\text{point-curve} = \text{envelope of line-curve}\}.$$

Unlike the example shown earlier in Fig. 7.1, where the line-curve image (i.e., the underlying envelope) is clear, in general the overlapping lines obscure portions of the resulting line-curve (see Exercise 1 below). In what follows, the intermediate steps are eliminated, resulting in a one-step equivalent $\{\text{point-curve}\} \rightarrow \{\text{point-curve}\}$ transformation having a natural generalization, as we see later on, to surfaces in \mathbb{R}^N .

Consider a planar curve c , which for the sake of generality is given implicitly by

$$c : F(x_1, x_2) = 0. \tag{7.6}$$

Rather than plot lines for the images of its points, the better way is to deal directly with the tangents at each point and convert them via (7.4) to points. Then, as illustrated on the right part of Fig. 7.10, determine the relation $\bar{c} : f(x, y) = 0$. This is not only a tremendous labor-saving device, but it also avoids completely plotting

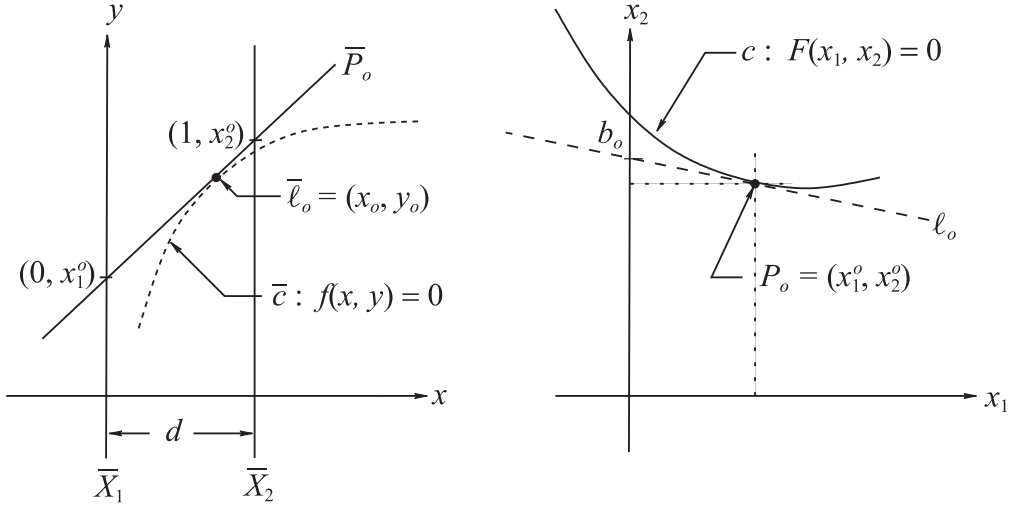


Figure 7.10. Obtaining the *point-curve* \bar{c} directly from the *point-curve* c .

and working with the line-curves. Proceeding, we obtain the total differential of F and the slope, i.e.,

$$dF = F_{x_1} dx_1 + F_{x_2} dx_2 = 0 \Rightarrow \frac{dx_2}{dx_1} = -\frac{F_{x_1}}{F_{x_2}}.$$

For a point (x_1, x_2) on the the tangent ℓ_o at the point $P_o = (x_1^o, x_2^o) \in c$, its slope is given by

$$m_o = -F_{x_1}^o / F_{x_2}^o = \frac{(x_2 - x_2^o)}{(x_1 - x_1^o)},$$

and the intercept is found by setting $x_1 = 0$ above, with the x_2 axis

$$b_o = \frac{x_2^o F_{x_2}^o + x_1^o F_{x_1}^o}{F_{x_2}^o}.$$

Substituting in (7.3) yields the point coordinates

$$x = d \frac{\partial F / \partial x_2}{(\partial F / \partial x_1 + \partial F / \partial x_2)}, \quad y = \frac{(x_1 \partial F / \partial x_1 + x_2 \partial F / \partial x_2)}{(\partial F / \partial x_1 + \partial F / \partial x_2)}, \quad (7.7)$$

where d is the interaxis distance and the superscript o is dropped, since the result is true for any point (x_1, x_2) . Ideal points of \bar{c} occur at the zeros of the denominator.

The smoothness of c ensures the continuity of \bar{c} . The result, (7.7), is known as the “*Inselberg transformation*” [41], [42].

Tracing the analogous steps for the transformation in the reverse direction $f(x, y) = 0 \rightarrow F(x_1, x_2) = 0$, we start with

$$df = f_x dx + f_y dy = 0 \Rightarrow \frac{dy}{dx} = -\frac{f_x}{f_y}.$$

The line \bar{P}_o that is tangent to the curve \bar{c} (see the left part of Fig. 7.10) at the point $\bar{\ell} = (x_o, y_o)$ has the equation

$$\bar{P}_o : -\frac{f_x^o}{f_y^o} = \frac{y - y_o}{x - x_o}.$$

Since it represents the point $P_o = (x_1^o, x_2^o)$, the intersections with the \bar{X}_1, \bar{X}_2 axes are at x_1^o and x_2^o respectively. In turn, these are obtained from the equation at $x = 0$ and $x = d$, providing

$$x_1 = \frac{x f_x + y f_y}{f_y}, \quad x_2 = \frac{(x - d) f_x + y f_y}{f_y}, \quad (7.8)$$

where again, reference to the original point (x_1^o, x_2^o) is dropped. In this way we have obtained the transformation:

$$c : \text{point-curve}(x_1 x_2 \text{ plane}) \leftrightarrow \bar{c} : \text{point-curve}(xy \text{ plane}),$$

where we have forsaken the messy intermediate step of having to deal with the line-curve. When the original point-curve is given explicitly by $x_2 = g(x_1)$, then (7.7) reduces to

$$x = \frac{d}{1 - g'(x_1)}, \quad y = \frac{x_2 + x_1 g'(x_1)}{1 - g'(x_1)}. \quad (7.9)$$

The value $d = 1$ is used throughout this chapter. The significance of varying d is better appreciated in Chapter 9, on surfaces. From now on, the images are shown as point-curve images, and occasionally, when there are no overlapping problems, as line-curves.

Exercises

1. Plot the line-curve corresponding to $x_2 = \sin x_1$. Hint: the inflection-point \leftrightarrow cusp duality is helpful in understanding the curve.
2. Find and plot the image of $x_2 = \sin x_1$ directly as a point-curve using (7.9).

3. Derive the analogue of (7.8) for $x_2 = g(x_1)$.
4. From (7.9), find the image of $g(x_1) = a_0 x_1^n + x_1$.
5. Derive equations (7.7) and (7.8) directly as the envelope of the family of lines corresponding to the points of the curve c given by (7.6).
6. Write a program that uses the above results to plot the point-curves \bar{c} for both implicitly and explicitly defined point-curves c using the above results. This can be a nice project topic.

7.4 ♣ FT-4 Curve Plotting

The image of a piecewise smooth curve can be computed and plotted via (7.7). Qualitatively, we can learn quite a bit from some elementary considerations of the duality by reviewing Fig. 7.11, which we saw earlier in Chapter 3, helping us sketch the curve's image. The portfolio of accompanying curves and their images, chosen to hone our experience, is drawn using a “curve-plotter” based on MATLAB.³¹ Due to the notational limitations of MATLAB, the bars “ $\bar{}$ ” are not included in the figures.

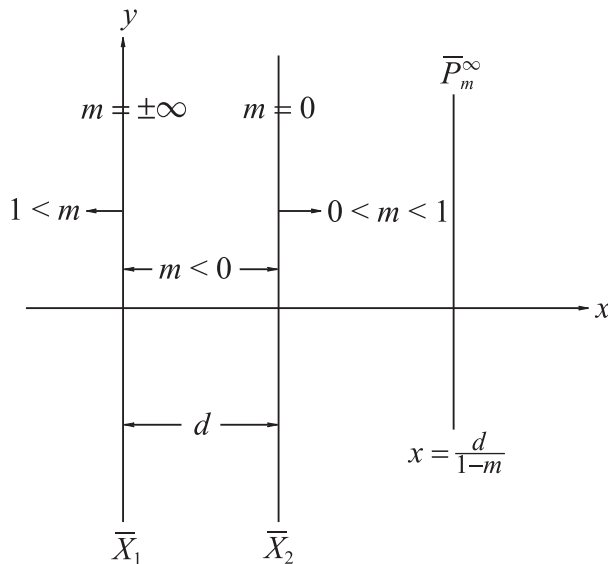


Figure 7.11. Horizontal position of $\bar{\ell}$ depends only on the slope m of ℓ .

³¹This beautiful curve-plotter (Exercise 6) was written by Tal Tzur and Gilat Schwartz for their course project in 2004.

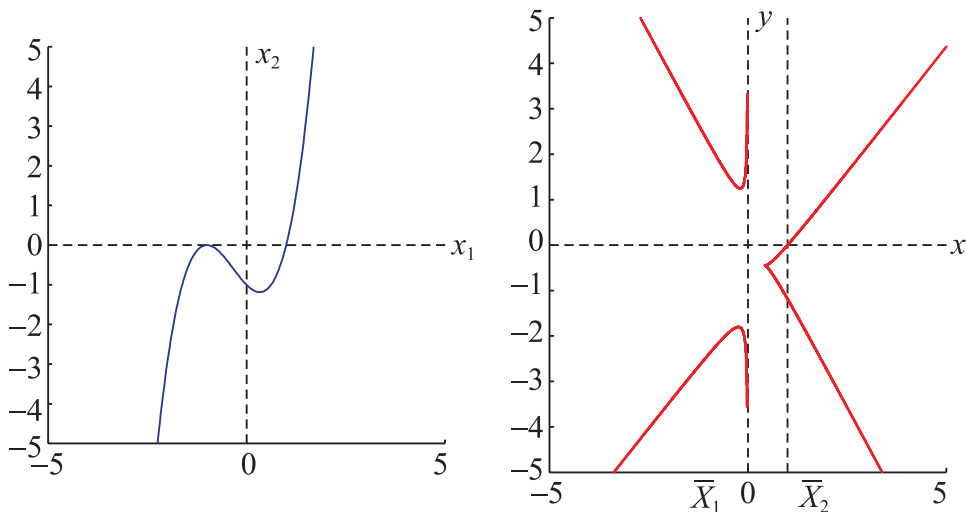


Figure 7.12. The algebraic curve $c : x_2 = x_1^3 + x_1^2 - x_1 - 1$ (left) has image (right) of degree 6.

In Section 7.6 it is pointed out that the image \bar{c} of an algebraic (i.e., described by a polynomial) curve c of degree n is also algebraic with degree n^* as given by the Plücker class formula (7.40), where s, d are the number of cusps and double-crossing points respectively. For the curve c on the left of Fig. 7.12, $n = 3$, $d = 0, s = 0$, hence for the image curve (right), we have $n^* = 6$. The analysis is facilitated by Fig. 7.13. The curve c has slope $m = 1$ at the points A_L and A_R , causing the image curve \bar{c} to split; the points of tangency there are mapped to ideal points. The portion c_I of c between A_L and A_R has the the right branch of \bar{c} as its image. According to Theorem 7.2.1, the inflection point I is mapped into the cusp \bar{i} of \bar{c} in between the two axes, since its slope m_i is negative. On c_I , the curve's slope m satisfies $m_i \leq m < 1$, and for this reason, \bar{c}_I opens to the right, intersecting \bar{X}_2 at two points: the images of the tangents at the extrema where $m = 0$. The curve's intersection with the \bar{X}_1 axis at two points is due to the zero points at the extrema; the higher intersection corresponds to M_a . By Corollary 7.1.4, the left portion of c_I , being cd , maps into the upper portion of the of \bar{c}_I , which is cu and approaches \bar{A}_L asymptotically. Similarly, the right portion of c_I , being cu , maps into the lower portion, which is cd , approaching \bar{A}_R asymptotically. The left portion of \bar{c} approaches the \bar{X}_1 axis asymptotically as $|x_1| \rightarrow \infty$, and the curve's slope m goes to ∞ . The upper portion of the left branch is cu , being the image of the portion $-\infty < x_1 < x_1(A_L)$, which is also cu . Note that the symmetry of c with respect

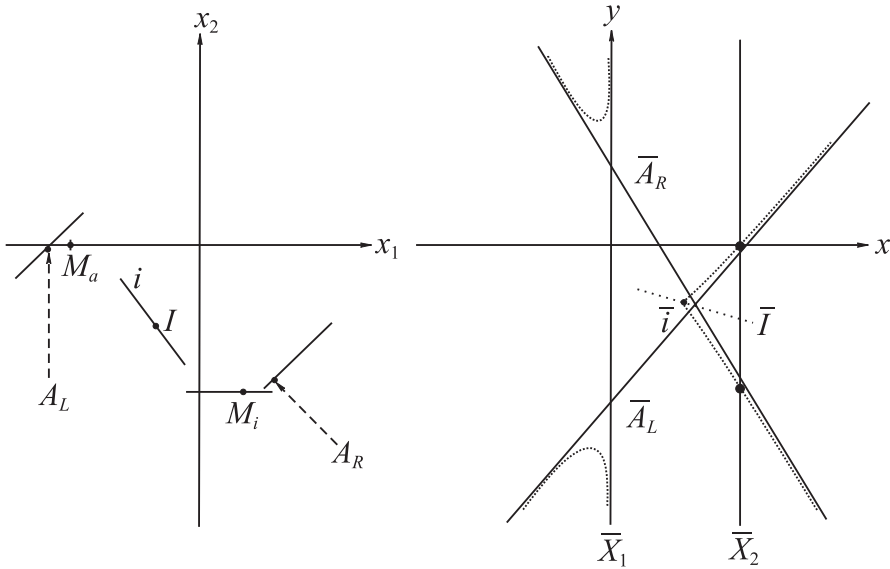


Figure 7.13. Analysis of the image curve.

The portions of the original curve c above are specified in terms of the slope (left) at the cubic curves' important points.

to the tangent i at I is transformed by symmetry with respect to the line \bar{i} through the cusp \bar{i} , the point image of the tangent i (see exercises below).

The curve c shown in Fig. 7.14 is prescribed by an implicit polynomial with $n = 3, s = d = 0$. Again \bar{c} has degree $n^* = 6$ but with two cusps stemming from the 2 ip of c . The image of the right branch \bar{c}_R of c is the portion containing the cusps. The slope m goes to ∞ as $x_1 \rightarrow \infty$ with \bar{c}_R approaching the \bar{X}_1 axis asymptotically. The (two) points of c where the slope m is equal to 1 are on the oval, splitting its image to the hyperbola-like part. Proceeding with another curve c (left) (Fig. 7.15) also prescribed by an implicit polynomial, again $n = 3, s = d = 0$, and the image curve \bar{c} has degree $n^* = 6$. There are two points of c where $m = 1$ responsible for the split of \bar{c} . The cusp to the right of the \bar{X}_2 axis is the image of the lower ip of c where the $0 < m < 1$, and also, since $m \rightarrow \infty$ as $x_1 \rightarrow \infty$, this part of \bar{c} approaches the \bar{X}_1 axis asymptotically. In the next example, shown in Fig. 7.16, the image curve has degree $n^* = 4$, since $n = 3, d = 1, s = 0$. Both points of c where $m = 1$ are to the right of the double-crossing point, with the part c_R to the right of these points mapped into the upper portion of c . The remaining part c_L of c containing the double point is mapped into two branches, each approaching asymptotically the negative \bar{X}_1 axis. The two tangents at the double point map into two points: the one on \bar{X}_2 is the image of the horizontal tangent. These two points

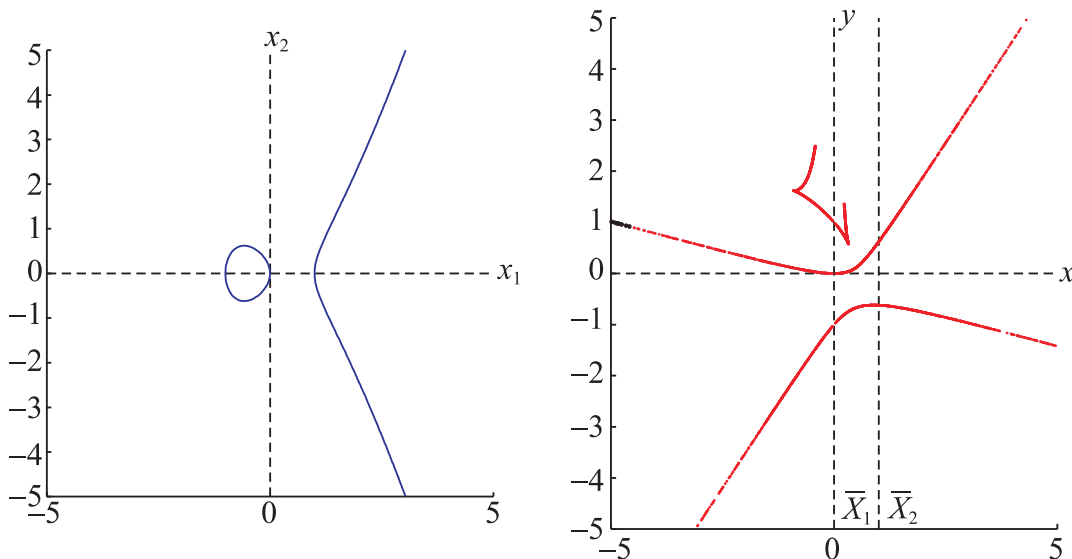


Figure 7.14. The image of the curve $c : F(x_1, x_2) = x_1^3 - x_1 - x_2^2 = 0$ has degree 6.

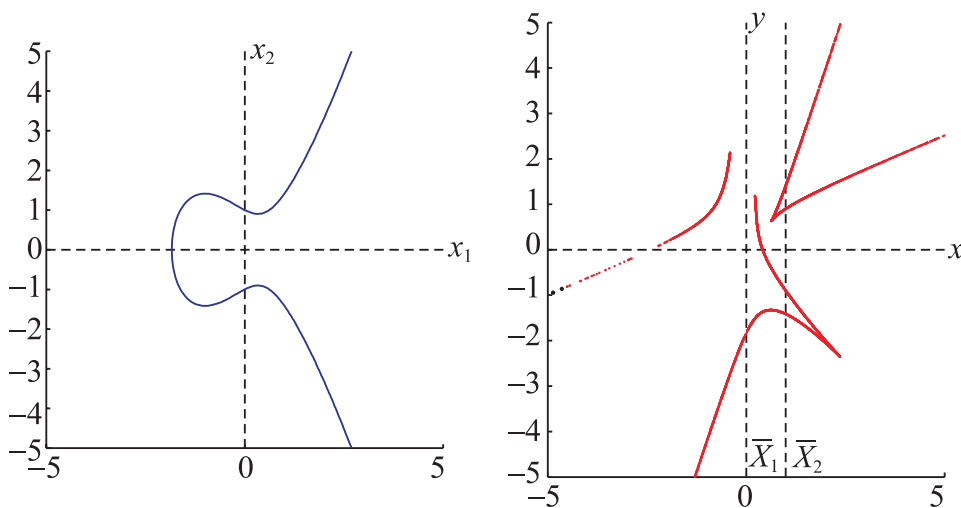


Figure 7.15. The image of the curve $c : F(x_1, x_2) = x_1^3 + x_1^2 - x_1 - x_2^2 = 0$ also has degree 6.

have the same tangent, the x axis. A tangent at two points of a curve is called a *bitangent*. The algebraic curve c shown in Fig. 7.17 is specified parametrically, and has $n = 3, s = 1, d = 0$, so that \bar{c} has degree $n^* = 3$. The two branches of c are tangent to the x_1 axis at the cusp, and therefore, it maps to the inflection point where

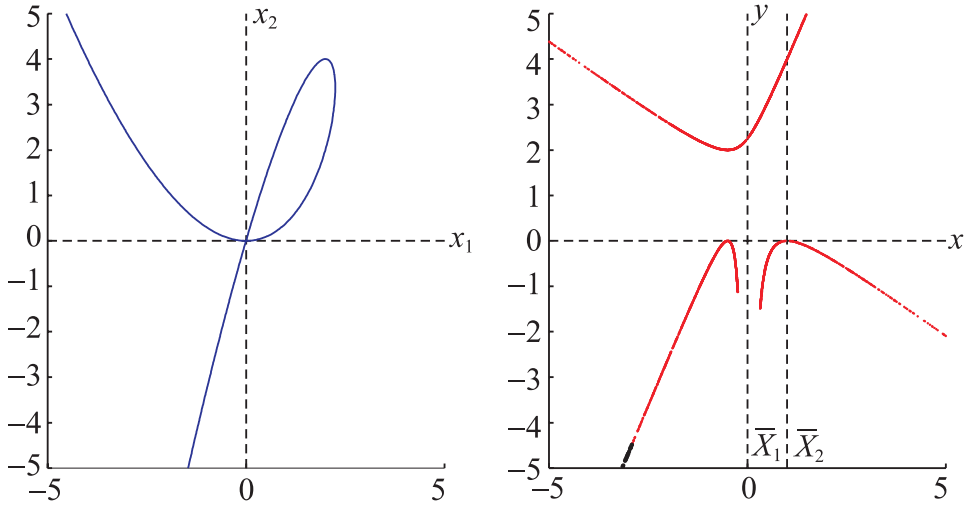


Figure 7.16. The image of the curve $c : F(x_1, x_2) = x_1^3 + x_1^2 - 3x_1x_2 = 0$ has degree 4.

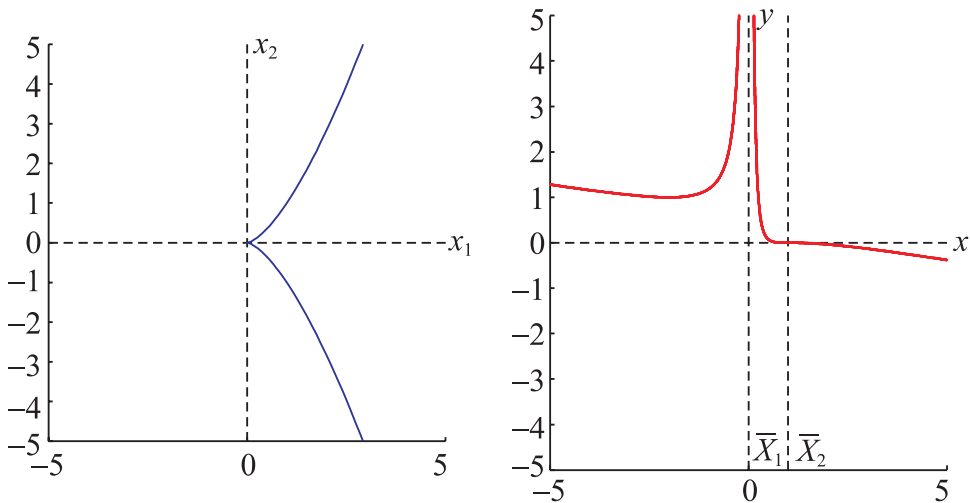


Figure 7.17. The image of the parametric polynomial $x(t) = t^2, x_2(t) = t^3$ has degree 3.

\bar{c} crosses the \bar{X}_2 axis. As in the previous examples, due to $m \rightarrow \infty$ as $x_1 \rightarrow \infty$, \bar{c} approaches the positive \bar{X}_1 axis asymptotically on either side. It is very easy to sketch the image of the exponential curve seen in Fig. 7.18. There is one point with slope $m = 1$ splitting the image curve \bar{c} into a part left of \bar{X}_1 , corresponding to the part on the right, where $1 < m < \infty$, and the left part of c , where $0 < m < 1$, whose

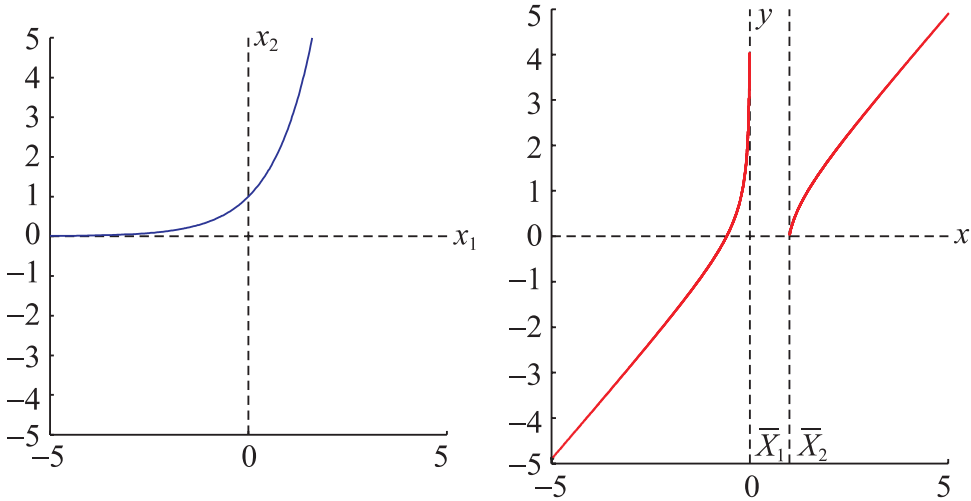


Figure 7.18. The image of the exponential curve $c : x_2 = e^{x_1}$. The image does not have a portion between the axes, since the curve does not have negative slopes.

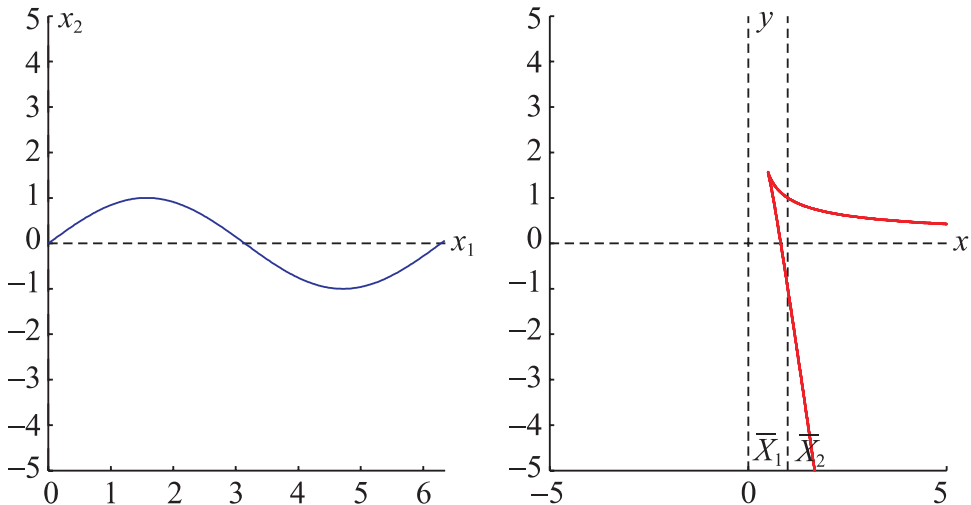


Figure 7.19. The image of $c : x_2 = \sin(x_1)$ in the interval $[0, 2\pi]$ is a nonoscillating curve.

image is to the right of the \bar{x}_2 axis. In the absence of negative slopes of c , \bar{c} does not have a portion between the \parallel -axes. It is interesting to trace the image of oscillatory curves starting with the trigonometric function $x_2 = \sin(x_1)$ plotted in Fig. 7.19 for the interval $x_1 \in [-\pi, 0]$. The ip at $x_1 = \pi$ maps into the cusp that is in between

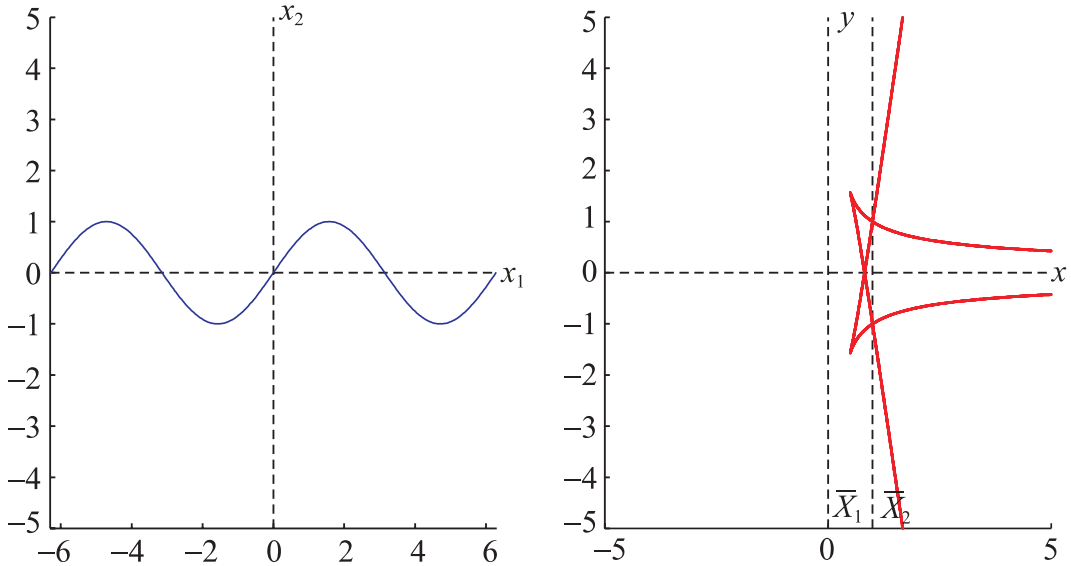


Figure 7.20. The image of $c : x_2 = \sin(x_1)$ for $x_1 \in [-2\pi, 2\pi]$.

It is symmetric about the x axis. The ip at $x_1 = \pm\pi$ are mapped into the two cusps, and the ip at the origin having slope $m = 1$ is mapped to the ideal point along the x axis.

the axes, since the tangents' slopes m are negative in its vicinity. In the remainder, the slope satisfies $0 < m(x_1) = \cos(x_1) \leq 1$, and hence \bar{c} opens to the right, the upper portion approaching the x axis asymptotically due to the ideal point from $x_1 = 0$ where the slope is 1. The graph for the interval $x_1 \in [0, 2\pi]$ is the mirror image of the curve above (Fig. 7.20), since the slopes satisfy $m(x_1) = \cos(-x_1)$. Altogether, then, the ip at $x_1 = \pm\pi$ with $m = 1$ are mapped into the two cusps with $x = 1/2$ and the ip at the origin, having slope $m = 1$, is mapped to the ideal point along the x axis. The mirror symmetry is preserved in the image of the curve $c : x_2 = x_1 \cos(x_1)$ in Fig. 7.21 and for analogous reasons.

7.4.1 Space-Curves

Curves and their images are an essential component in the representation of surfaces (Chapter 9) in terms of planar regions whose boundaries are the images of certain *space-curves*. It will be seen that via the wonderful properties of the superplanes, the boundaries can be obtained from images of *plane-curves* as studied in this chapter. Still, a few words about space-curves are obligatory. Let us start with a convenient parametric definition with sc denoting a space-curve:

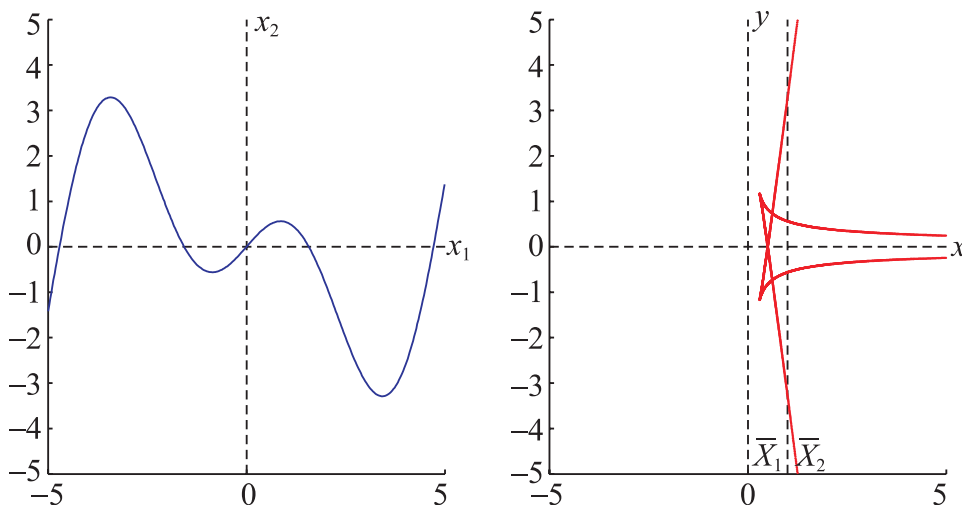


Figure 7.21. The image of the oscillatory curve $x_2 = x_1 \cos(x_1)$. It is a nonoscillating curve symmetric with respect to the x axis.

$$sc : x_i = x_i(s), \quad s \in I \subset \mathbb{R}, \quad i = 1, 2, 3, \quad (7.10)$$

where I is an interval and the parameter s is often taken as the arc length. As for plane-curves, the *tangent* t at a point P is found as the limiting chord joining P to an adjoining point. Two other adjoining points Q, R together with P form a plane, which in the limit as $(Q, R) \rightarrow P$ is the *osculating plane*—a rather endearing appellation—containing the tangent t . Within the osculating plane, the local center of curvature of sc at P is found as for plane-curves and the *principal normal* \mathbf{n} at P directed towards the center is chosen. It is called *principal* to distinguish it from the infinity of normals on the plane orthogonal to t at P , or just *normal* when the context is clear. The directed tangent \mathbf{t} is chosen so that together with \mathbf{n} as axes they form a right-handed system. Next, the *binormal* \mathbf{b} is included orthogonal to the osculating plane and directed so that $\mathbf{t} \rightarrow \mathbf{n} \rightarrow \mathbf{b}$ has the same sense as $x_1 \rightarrow x_2 \rightarrow x_3$, i.e.,

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}, \quad (7.11)$$

the three usually taken as unit vectors. The rate of change

$$\frac{d\mathbf{b}}{ds} = -\tau \mathbf{n} \quad (7.12)$$

introduces the *torsion* τ , which measures the “nonplanarity,” that is, the change of the osculating plane along the curve with $\tau = 0$ for plane curves. The derivatives

of \mathbf{b} , \mathbf{t} , \mathbf{n} provide the important equations of *Frenet* relating the three unit vectors to the two fundamental properties: torsion and curvature. Prescribing the curvature and torsion as functions of the arc length completely determines a space-curve up to its position and orientation. We leave (see Exercise 13) this interesting topic after this very short introduction in order to return to plane curves for the remainder of the chapter.

Exercises

1. Prove that the symmetry about the tangent i of the ip I corresponds to the symmetry about the line \bar{I} through \bar{i} , or find portions of \bar{c} where the symmetry fails.
2. What points correspond to the max and min of the left-hand branch in Fig. 7.12?
3. Is the *double-point* \leftrightarrow *bitangent* duality seen in Fig. 7.14 true in general? Prove or give a counterexample.
4. For the curve in Fig. 7.14, provide an analysis of its image and draw a figure of the important points with the tangents similar to Fig. 7.13.
5. What portions of \bar{c} in Fig. 7.14 are symmetric about a line and why?
6. Analyze the image of the curve in Fig. 7.15, sketching a figure similar to Fig. 7.13.
7. Is there an algebraic curve of degree 3 whose image has degree 5? Give an example of an algebraic curve of any degree having image with degree 5.
8. Rotate the exponential curve in Fig. 7.18 to obtain an image having points for all values of x .
9. In Fig. 7.21, mark the portions of c and their corresponding parts in the image \bar{c} .
10. Prove that the image \bar{c} in Fig. 7.19 consists of monotone increasing and monotone decreasing portions, and find the portions of c corresponding to each.
11. Let $c : x_2 = f(x_1)$ be an odd function (i.e., $f(x_1) = -f(x_1)$). Is \bar{c} necessarily symmetric about the x -axis?
12. For a general curve c given by (7.6), the image of the tangent at each point $P \in c$ is the point given by (7.7). Find
 - (a) the image of the *normal* at P .
 - (b) the formula for the curvature of the image curve \bar{c} ; *hard*.
13. Space curves
 - (a) Find the image \overline{cs} of a smooth *space-curve* $cs : x_i = x_i(s) \subset \mathbb{R}^3$, $i = 1, 2, 3$. Hint: recall from Fig. 7.10 that the image of a curve is

Table 7.1. Equalities between invariants of Plücker curves and their duals.

Point(s) on curve $c \rightarrow$	map into points of the curve $\bar{c} \Rightarrow$	relation
The two points of c on a bitangent	map into a double point of \bar{c}	$b = d^*$
A double point of c	maps into two points on a bitangent of \bar{c}	$d = b^*$
An inflection point of c	maps into a cusp of \bar{c}	$ip = s^*$
A cusp of c	maps into an inflection point of \bar{c}	$s = ip^*$

constructed from the tangents at each point. For a space-curve the tangent at each point is a line ℓ in \mathbb{R}^3 and is represented by two points $\bar{\ell}_{12}, \bar{\ell}_{23}$, rather than one point as for \mathbb{R}^2 . Hence the image \overline{cs} consists of two curves, say \overline{cs}_{12} and \overline{cs}_{23} , in the xy plane.

- (b) For the space-curve c above, using the property that the three points $\bar{\ell}_{12}, \bar{\ell}_{23}, \bar{\ell}_{13}$ are collinear, how can the curve \overline{cs}_{12} be determined from the other two curves?
- (c) Obtain generalizations of the dualities listed in Table 7.1.
- (d) Find the image \overline{cs} of a smooth *space-curve* $cs : x_i = x_i(s) \subset \mathbb{R}^N$, $i = 1, \dots, N$, consisting of $N - 1$ curves in the xy plane and generalize the dualities in Table 7.1.

This is an excellent project topic.

7.5 FT-5 Transforms of Conics

The conic sections are described by the quadratic function

$$\begin{aligned}
 F(x_1, x_2) &= A_1 x_1^2 + 2A_4 x_1 x_2 + A_2 x_2^2 + 2A_5 x_1 + 2A_6 x_2 + A_3 = 0 \\
 &= (x_1, x_2, 1) \begin{pmatrix} A_1 & A_4 & A_5 \\ A_4 & A_2 & A_6 \\ A_5 & A_6 & A_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}, \quad (7.13)
 \end{aligned}$$

where the type of conic is determined by the sign of the *discriminant* $\Delta = (A_4^2 - A_1 A_2)$. The coefficient matrix is denoted by A , and its determinant, which plays an important role in the development, is

$$\det A = A_3(A_1 A_2 - A_4^2) - A_1 A_6^2 - A_2 A_5^2 + 2A_4 A_5 A_6. \quad (7.14)$$

In this section it is first shown that conics \leftrightarrow conics, and then criteria for the determination of specific conic images are developed.

7.5.1 ** Proving the Duality Conics \leftrightarrow Conics

For conics, using the identity that for a polynomial F of degree n , $F(\mathbf{x}) = 0 \Rightarrow \nabla F \cdot \mathbf{x} = \nabla F \cdot \mathbf{x} - nF$, the second expression being linear, (7.7) becomes

$$\begin{aligned} x &= d \frac{A_4x_1 + A_2x_2 + A_6}{[(A_1 + A_4)x_1 + (A_2 + A_4)x_2 + (A_5 + A_6)]}, \\ y &= -\frac{A_5x_1 + A_6x_2 + A_3}{[(A_1 + A_4)x_1 + (A_2 + A_4)x_2 + (A_5 + A_6)]}. \end{aligned} \quad (7.15)$$

Note that the interaxis distance d enters as a multiplier only in the x coordinate due to the *line* \rightarrow *point* mapping. These are *Möbius*³² transformations, which form a group (see any good book in modern algebra or [10]). This observation enables substantial simplifications of the earlier treatment of conics and their transforms (see [41] and [42]).³³ The *inverse*, expressing x_1 and x_2 in terms of x and y , is a Möbius transformation of the form

$$x_1 = \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, \quad x_2 = \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}}, \quad (7.16)$$

which can also be written as

$$\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} / (a_{31}x + a_{32}y + a_{33}), \quad \alpha = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \quad (7.17)$$

It is also known that the Möbius transformations are homomorphic to the group of nonsingular matrices of appropriate size. Without entering into unnecessary details, the Möbius transformation

$$u = \frac{b_{11}x + b_{12}y + b_{13}}{b_{31}x + b_{32}y + b_{33}}, \quad v = \frac{b_{21}x + b_{22}y + b_{23}}{b_{31}x + b_{32}y + b_{33}}, \quad (7.18)$$

corresponds to the matrix $M_1 = [b_{ij}]$, where it is required that $\det(M_1) = 1$, normalizing if necessary. That does not pose a problem if M_1 is nonsingular, since the numerators and denominators of the fractions in (7.18) can be multiplied by the same constant without altering the transformation. In our case, the matrix corresponding to the Möbius transformation (7.15) is

³²Also called *linear rational* transformations.

³³The result *conics* \leftrightarrow *conics* was first proved by B. Dimsdale.

$$C' = \begin{pmatrix} dA_4 & dA_2 & dA_6 \\ -A_5 & -A_6 & -A_3 \\ (A_1 + A_4) & (A_2 + A_4) & (A_5 + A_6) \end{pmatrix} = \begin{pmatrix} 0 & d & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix} A = RA. \quad (7.19)$$

The matrix R is obtained by applying row operations on C' to achieve the indicated reduction. Noting that $\det(C') = -\det(A) = H$, we normalize by $C = C'/H = RA/H$ to obtain $\det(C) = 1$ so as to satisfy the requirement of the aforementioned homomorphism. This corresponds to multiplying both numerator and denominator of the Möbius transformation (7.15), or equivalently the elements of the corresponding matrix, by $-(\det A)^{-1/3}$. The coefficients, and in turn the corresponding matrix $\alpha = [a_{ij}]$, in (7.16), which is the inverse Möbius transformation of (7.15), are simply the elements of C^{-1} . In terms of the original coefficients,

$$(C')^{-1} = HC^{-1} = HA^{-1}R^{-1} = -\text{adj}(A)R^{-1} = \alpha, \quad (7.20)$$

where $\text{adj}(A) = [A_{ij}]^T$ is the adjoint, with the A_{ij} being the cofactors of the matrix A . Proceeding with the evaluation,

$$\begin{aligned} (C')^{-1} &= - \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} \begin{pmatrix} -1/d & 0 & 1 \\ 1/d & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} (A_{11} - A_{21})/d & A_{31} & -A_{11} \\ (A_{12} - A_{22})/d & A_{32} & -A_{12} \\ (A_{13} - A_{23})/d & A_{33} & -A_{13} \end{pmatrix}. \end{aligned} \quad (7.21)$$

The individual a_{ij} are

$$\left. \begin{aligned} a_{11} &= (A_{11} - A_{21})/d = [A_3(A_2 + A_4) - (A_6^2 + A_5A_6)]/d, \\ a_{12} &= A_{31} = A_4A_6 - A_2A_5, \\ a_{13} &= -A_{11} = A_6^2 - A_2A_3, \\ a_{21} &= (A_{12} - A_{22})/d = [(A_5^2 + A_5A_6) - A_3(A_1 + A_4)]/d, \\ a_{22} &= A_{32} = A_4A_5 - A_1A_6, \\ a_{23} &= -A_{12} = A_3A_4 - A_5A_6, \\ a_{31} &= (A_{13} - A_{23})/d = [A_6(A_1 + A_4) - A_5(A_2 + A_4)]/d, \\ a_{32} &= A_{33} = A_1A_2 - A_4^2, \\ a_{33} &= -A_{13} = A_2A_5 - A_4A_6. \end{aligned} \right\} \quad (7.22)$$

Recapitulating,

$$\begin{aligned} F(x_1, x_2) &= (x_1 \ x_2 \ 1)A \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \\ &= (x \ y \ 1)\alpha^T A\alpha \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \Big/ (a_{31}x + a_{32}y + a_{33})^2 = 0, \end{aligned} \quad (7.23)$$

for $(a_{31}x + a_{32}y + a_{33})^2 \neq 0$. Or

$$f(x, y) = H(x \ y \ 1)\mathbf{a} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0, \quad \mathbf{a} = \frac{1}{H}(\alpha^T A\alpha) = \begin{pmatrix} a_1 & a_4 & a_5 \\ a_4 & a_2 & a_6 \\ a_5 & a_6 & a_3 \end{pmatrix}. \quad (7.24)$$

Further,

$$\mathbf{a} = (R^{-1})^T (A^{-1})^T A A^{-1} R^{-1} H = (R^{-1})^T \alpha, \quad (7.25)$$

where $(A^{-1})^T = (A^T)^{-1} = A^{-1}$, since A is symmetric. Therefore,

$$\begin{aligned} \mathbf{a} &= \begin{pmatrix} a_1 & a_4 & a_3 \\ a_4 & a_2 & a_6 \\ a_5 & a_6 & a_3 \end{pmatrix} \\ &= \begin{pmatrix} -1/d & 1/d & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} (A_{11} - A_{21})/d & A_{31} & -A_{11} \\ (A_{12} - A_{22})/d & A_{32} & -A_{12} \\ (A_{13} - A_{23})/d & A_{33} & -A_{13} \end{pmatrix} \\ &= \begin{pmatrix} (2A_{12} - A_{22} - A_{11})/d^2 & (A_{23} - A_{13})/d & (A_{11} - A_{12})/d \\ (A_{23} - A_{13})/d & -A_{33} & A_{13} \\ (A_{11} - A_{12})/d & A_{13} & -A_{11} \end{pmatrix}. \end{aligned} \quad (7.26)$$

The reduction is due to the symmetry of A and hence of its cofactors, i.e., $A_{ij} = A_{ji}$, confirming that \mathbf{a} is symmetric. The individual a_i are

$$\left. \begin{aligned}
 a_1 &= [A_3(A_1 + A_2 + 2A_4) - (A_5 + A_6)^2]/d^2, \\
 a_2 &= A_1A_2 - A_4^2, \\
 a_3 &= A_2A_3 - A_6^2, \\
 a_4 &= [A_{13} - A_{23}]/d = [A_6(A_1 + A_4) - A_5(A_2 + A_4)]/d, \\
 a_5 &= [A_6^2 + A_5A_6 - A_3(A_2 + A_4)]/d, \\
 a_6 &= A_2A_5 - A_4A_6.
 \end{aligned} \right\} \quad (7.27)$$

The corresponding results for the inverse transform, i.e.,

$$\text{conics in the } xy \text{ plane} \mapsto \text{conics in the } x_1x_2 \text{ plane}$$

are left as an exercise. The general results obtained here enable the study and complete classification of the conic transforms. The transforms of specific conics are obtained from the criteria developed next.

7.5.2 FT-6 Classification of the Conic Transforms

Projectively, conics are ellipses having 0 (ellipse), 1 (parabola), or 2 (hyperbola) ideal points. The conic's ideal point(s) can be found by taking $\lim_{x_1 \rightarrow \infty} (x_2/x_1) = m$, where m is a constant (i.e., the direction of the ideal point), from (7.13), yielding the relation

$$A_2m^2 + 2A_4m + A_1 = 0. \quad (7.28)$$

To include the vertical direction (i.e., where the slope is $\pm = \infty$), evaluating the reciprocal $\lim_{x_2 \rightarrow \infty} (x_1/x_2) = M$ provides

$$A_1M^2 + 2A_4M + A_2 = 0. \quad (7.29)$$

The conic's *indicator* I is defined as the sum of the coefficients

$$I = A_1 + A_2 + 2A_4. \quad (7.30)$$

The conic transforms are given by

$$f(x, y) = a_1x^2 + 2a_4xy + a_2y^2 + 2a_5x + 2a_6y + a_3 = 0, \quad (7.31)$$

the conic type being determined from the sign of

$$\delta = a_4^2 - a_1a_2 = H(A_1 + A_2 + 2A_4) = HI. \quad (7.32)$$

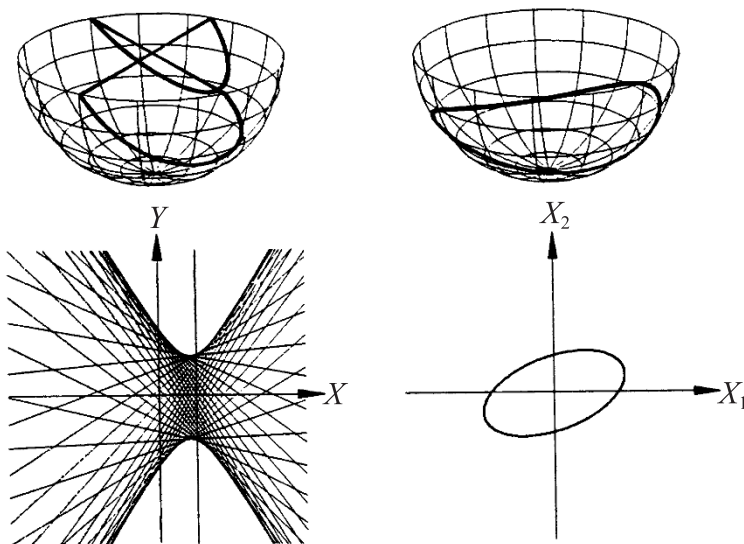


Figure 7.22. Ellipses always map into hyperbolas.

Each asymptote is the image of a point where the tangent has slope 1.

When $H < 0$, the conic has no real points.³⁴ When $H = 0$, the conic is a point. The other two *degenerate* cases occur for conics consisting either of a single line or a pair of intersecting lines. In the ensuing it is assumed that $H > 0$, which fixes the signs of the quadratic's coefficients. In view of (7.32), the indicator I together with the sign of Δ tells the story. For a particular conic c , *only* the coefficients A_1 , A_2 , and A_4 determine the type of conic \bar{c} . *So rotations of c , as reflecte by A_4 , may change the type of its transform, while translations, which are specifie by the linear and constant terms cannot.* The analysis below is listed by conic types.

Ellipse, $\Delta < 0$. Then $A_1 A_2 > A_4^2 \geq 0$; hence A_1 and A_2 have the same sign. If they are positive, then $A_1 > A_4^2/A_2$ and

$$I = (A_1 + A_2 + 2A_4) > (A_4^2 + A_2^2 + 2A_4 A_2)/A_2 = (A_2 + A_4)^2/A_2 \geq 0,$$

so the transform is a *hyperbola* (Fig. 7.22), since $\delta > 0$. The same inequality holds if A_1 and A_2 are both negative. For an ellipse, of course, the m_i are complex conjugates.

³⁴The quadratic (7.13) may be solved for x_2 in terms of x_1 and the coefficients A_i . There are no real values of x_2 if the discriminant is negative. This provides a quadratic inequality in x_1 whose discriminant turns out to be H .

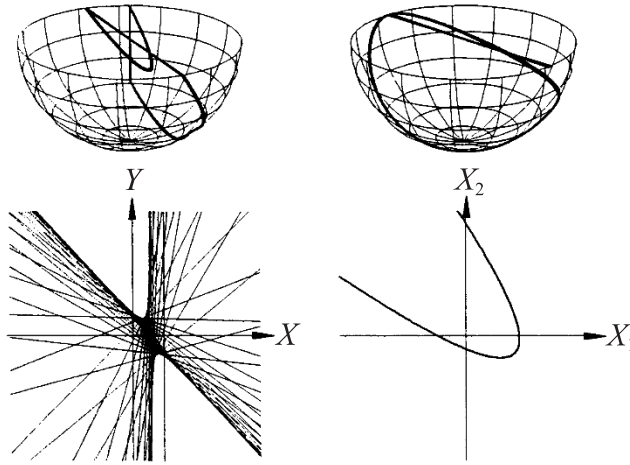


Figure 7.23. A parabola whose ideal point does not have direction with slope 1. It *always* transforms to a hyperbola with a vertical asymptote. The other asymptote is the image of the point where the parabola has tangent with slope 1.

Parabola, $\Delta = 0$. If any one of A_1, A_2, A_4 is 0, then all three are 0, and (7.13) does not describe a quadratic function. So $A_2 \neq 0$ and $A_1 = A_4^2/A_2$. Then

$$\left. \begin{aligned} (A_1 + A_2 + 2A_4) &= (A_2 + A_4)^2/A_2, \\ (A_1 + A_4) &= (A_2 + A_4)A_4/A_2, \\ H &= A_1A_6^2 + A_2A_5^2 - 2A_4A_5A_6 = (A_4A_6 - A_2A_5)^2/A_2, \\ a_4 &= (A_2 + A_4)(A_4A_6 - A_2A_5)/A_2, \\ a_1 &= [A_3(A_2 + A_4)^2 - A_2(A_5 + A_6)^2]/A_2. \end{aligned} \right\} \quad (7.33)$$

Hence if $I \neq 0$, then $A_2 \neq -A_4$, $\delta = HI > 0$, and the transform is a *hyperbola*, as shown in Fig. 7.23. Otherwise, when $I = 0$, from (7.34), the parabola's ideal point has slope 1, and the transform is a *parabola*. This *self-dual* case is shown in Fig. 7.24.

Hyperbola, $\Delta > 0$. If $I < 0$, then the transform is an *ellipse*, as shown in Fig. 7.25. This is the dual of the *ellipse* \mapsto *hyperbola* transform.

Otherwise, if $I = 0$ (see (7.28)), then one of the hyperbola's ideal points has slope 1 and the transform is a *parabola* (Fig. 7.26). This is the dual of the *parabola* \mapsto *hyperbola* transform.

Lastly, if $I > 0$, the transform is a *hyperbola*. This *self-dual* case is shown in Fig. 7.27.

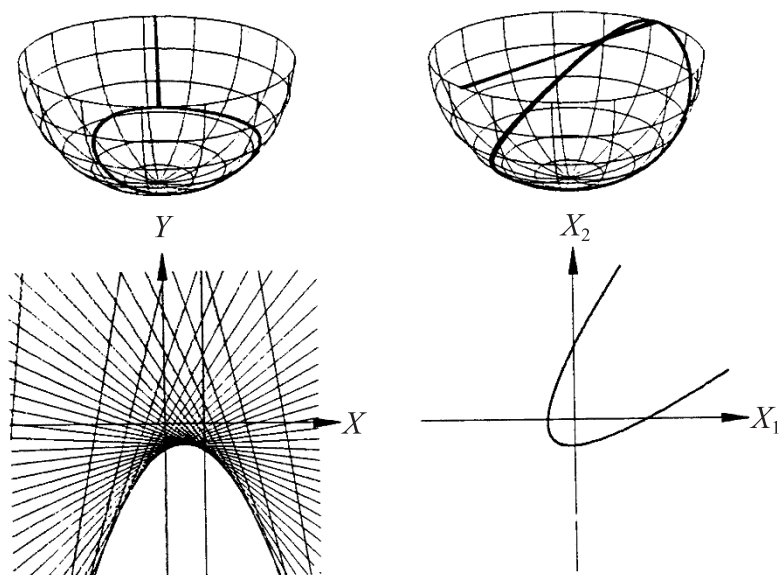


Figure 7.24. A parabola whose ideal point has slope 1 transforms to a parabola; self-dual.

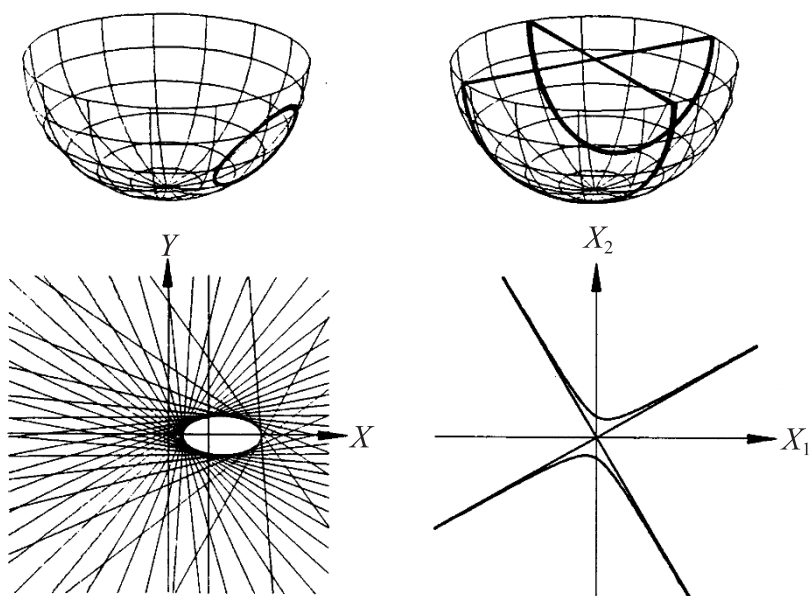


Figure 7.25. Hyperbola to ellipse; dual of the case shown in Fig. 7.22.

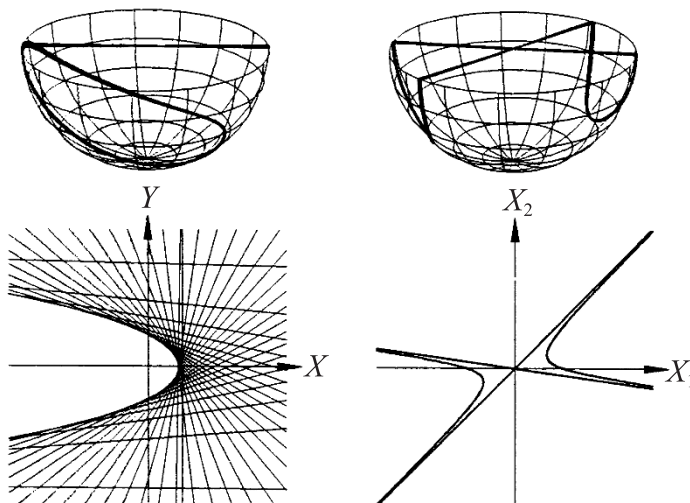


Figure 7.26. If the asymptote has slope 1, then the hyperbola \rightarrow parabola; dual of Fig. 7.23.

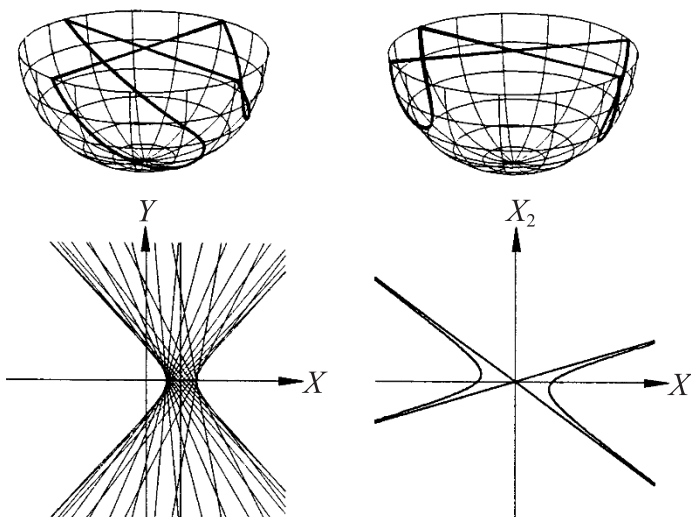


Figure 7.27. Hyperbola to hyperbola; self-dual case.

Recasting the indicator I in terms of the conic's ideal points provides not only an intuitive way of understanding the transforms but also, as shown in the next section, a pointer to generalizations. For $A_2 \neq 0$ let m_i , $i = 1, 2$, be the roots of (7.28) with

$m_1 \leq m_2$ when they are real. From the coefficients expressed in terms of the roots by $A_1 = A_2 m_1 m_2$, $2A_4 = -A_2(m_1 + m_2)$,

$$I = A_2(1 - m_1)(1 - m_2), \quad (7.34)$$

with one of the m_i equal to 0 when $A_1 = 0$. For $A_2 = 0$ the roots of (7.29) are $M_1 = 1/m_1 = A_4/A_1$, the *only* one also obtained from (7.28), and $M_2 = 0$ (i.e., $m_2 = \pm\infty$). Finally, when both $A_1 = A_2$ are zero, the two roots, obtained one from each of the equations, are $m_1 = 0$ and $M_2 = 0$ ($m_2 = \pm\infty$).

Whence Come the Hyperbolas?

We have seen that hyperbolas occur as images of ellipses, parabolas, and also hyperbolas. Is there a way to decide from the hyperbola what kind of conic it is the image of? To clarify this, it is necessary to trace the image of interior points. The interior points of an ellipse are mapped via the 2-D *point* \longleftrightarrow *line* into lines that are between the hyperbola's two branches. This is also true for the two cases (see exercises below). Even without this observation, the easiest case is for a hyperbola that has a vertical asymptote. From the foregoing analysis, this can only be the image of a parabola, where the vertical asymptote represents the parabola's ideal point. This is seen well in Fig. 7.23, where the parabola opens in the direction with slope $m = -1$ and the vertical asymptote is at $x = 1/2$. The other asymptote is, of course, determined by the point where the parabola has a tangent with slope $m = 1$ and is discussed below, providing (7.38).

To distinguish the remaining two cases, it is worth scrutinizing the left portions of Figures 7.22 and 7.27. The essential difference and clue is the horizontal separation along the x axis of the two left and right branches in Fig. 7.27, as contrasted with the vertical separation, along the y axis, of the top and bottom branches. Note that the hyperbola on the right side of Fig. 7.27 contains the ideal points whose directions are between the upper left asymptote and the upper right asymptote. That is all the directions of the lines that cannot be tangents to the hyperbola. All these ideal points are mapped into the vertical lines between the two branches and hence the horizontal separation. Of course, there are also regular points between the branches of the original hyperbola, and these are mapped into nonvertical lines between the branches of the image hyperbola. By contrast, the ellipse has no ideal points, and hence there cannot be any horizontal separation between the branches of its image. Rather, all interior points appear as lines between the two branches.

Alternative Analysis Based on the Indicator

Now let us revisit the results above in view of (7.34).

- When the m_i are complex conjugates, the conic is an ellipse with $I > 0$.
- Otherwise, when there is a single real root m_1 , the conic is a parabola, and the transform is
 - a hyperbola with one vertical asymptote when $m_i \neq 1$, and
 - a parabola when $m_i = 1$.
- Finally, when there are two distinct real roots, the conic is a hyperbola whose asymptotes have slopes m_i . There are two such hyperbolas: one whose tangents have slope $m \in [m_1, m_2]$ and the other with slope $m \notin [m_1, m_2]$. For easy reference we refer to the two hyperbolas having the same asymptotes as *conjugate*. The conic's transform is determined by whether $1 \in [m_1, m_2]$. Recalling (7.4):
 - when $m_1 < 1 < m_2$, the points of the conic's transform have $x = d/(1 - m)$, whose range is the *finite* interval $[1/(1 - m_2), 1/(1 - m_1)]$, hence the transform must be an ellipse, $I < 0$.
 - For the *conjugate* hyperbola whose tangents have slopes $m \notin (m_1, m_2)$ (i.e., two semi-infinite intervals), the transform must be a hyperbola, $I > 0$.
 - For $m_1 < 1 = m_2$, one hyperbola has tangents with slope $m \in (m_1, 1)$. The conjugate hyperbola's tangents have slope $m \notin (m_1, 1)$ as shown in Fig. 7.26. Since $I = 0$, the images of both hyperbolas are parabolas tangent at their vertices and the reflection of each other.

7.5.3 ** The Conic Transform's Ideal Points

It is instructive to investigate the conic transforms from their ideal points. Again we have to consider two limits to obtain the corresponding equations. To wit, $\lim_{x \rightarrow \infty} y/x$ in (9.21) yields

$$a_2 R^2 + 2a_4 R + a_1 = 0 \Rightarrow R = \frac{-a_4 \pm \sqrt{HI}}{a_2}, \quad (7.35)$$

and $\lim_{y \rightarrow \infty} (x/y) = r$ for

$$a_1 r^2 + 2a_4 r + a_2 = 0 \Rightarrow r = \frac{-a_4 \pm \sqrt{HI}}{a_1}. \quad (7.36)$$

Let us now investigate the conic transform's ideal points methodically.

- When $I < 0$, the conic is a hyperbola with $m_1 < 1 < m_2$, whose transform, being an *ellipse*, has no ideal points.
- For $I = 0$, we have $A_1 + A_4 = -(A_2 + A_4)$, and the conic's transform is a *parabola*. If the original conic is a *parabola*, then $-\Delta = a_2 = 0$, $r = -a_4/a_1$, and from (7.33), $A_2 + A_4 = 0$, $a_4 = 0 \Rightarrow r = 0$. Hence the transform's ideal point is in the vertical direction as seen in Fig. 7.24. Otherwise, the original conic is a *hyperbola*. Then

$$r = -\frac{A_2 + A_4}{A_5 + A_6}, \quad (7.37)$$

and the transform's ideal point has slope $1/r = R$. Note that this is due to $\delta = 0 \Rightarrow a_4/a_1 = a_2/a_4$.

- Finally, for $I > 0$, r has two real values, and the transform is a *hyperbola*. If the original conic is an *ellipse*, the transform has asymptotes with finite slopes $1/r$, as shown in Fig. 7.22.

If it is a *parabola*, then the roots are $r = 0$, which accounts for the vertical asymptote, and

$$r = -2\frac{a_4}{a_1}, \quad (7.38)$$

which provides the other asymptote shown in Fig. 7.23.

If it is a *hyperbola*, this is self-dual, and the transform has asymptotes with finite slopes $1/r$, as seen in Fig. 7.27.

Observe that the transforms of a parabola *always* have an ideal point in the vertical direction ($r = 0$). In turn, this is because each parabola's ideal points are mapped into a vertical line by the basic point \leftrightarrow line. A hyperbola's two ideal points are also mapped to vertical lines, which turn out to be vertical *tangents* of the corresponding transform, i.e., the vertical tangents at the ellipses' endpoints in Fig. 7.25, the parabola's vertex in Fig. 7.26, and the two vertices of the hyperbola in Fig. 7.27.

To summarize, the transform's ideal points are either

- the images of the conic's ideal points $r = 0$ for parabolas as shown in Fig. 7.24 and 7.27, or
- the images of points where the conic has tangents with slope 1. They correspond to the real nonzero roots of (7.28).

** Eliminating the Conics' Linear Terms by Translation

The conic transform's ideal points are very sensitive to changes in the original conic's orientation. It is worthwhile to study the effects of translation. For this

reason, the linear terms in (7.13) are omitted (i.e., $A_5 = A_6 = 0$), since they can be eliminated with a translation. Then

$$\left. \begin{aligned} a_1 &= A_3(A_1 + A_2 + 2A_4), \\ a_2 &= A_1A_2 - A_4^2, \\ a_3 &= A_2A_3, \\ a_4 &= 0, \\ a_5 &= A_3(A_2 + A_4), \\ a_6 &= 0, \end{aligned} \right\} \quad (7.39)$$

with $\delta = HI$ and $H = A_3\Delta$. It is striking that the cross term vanishes ($a_4 = 0$). The resulting conic transforms are in *standard position* with their symmetry axes parallel to either the x or y axis. For illustration, the parabola in Fig. 7.23 is not centered, and its transform is not in standard position. The reason for the parabola's horizontal ideal point (i.e., $1/r = 0$) in Fig. 7.26 is seen from (7.37). This is due to the duality

$$\left\{ \begin{array}{l} \text{translation of a} \\ \text{point along a line} \end{array} \right\} \iff \left\{ \begin{array}{l} \text{rotation of a} \\ \text{line about a point} \end{array} \right\}.$$

This observation deserves further analysis, since there is a translation in the x but not in the y direction.

Note that all the transforms in the figures, except Fig. 7.23, are translated in the x direction. Further, the original conic is centered, and the resulting conic transform is not rotated from the standard position. A surprising observation is that only five of the six cases are found from the results above when the linear terms are eliminated. This is because when A_5 and A_6 are zero, a parabola's H vanishes and the two roots of (7.28) collapse to $M = 0$ (see (7.38)). For this reason, the parabola in Fig. 7.23 is carefully not centered in order to avoid this. Recall that the normalization of the determinant matrix's C' , (7.19), involved division by H , and hence the method breaks down for $H = 0$ (see exercises below).

Exercises

1. Using the theory of envelopes, find the transforms of the three different kinds of conics in
 - (a) standard position,
 - (b) translated in the x and separately in the y direction,
 - (c) with the cross term not zero.

2. What is the conic transform of the hyperbola $x_2 = 1/x_1$ and its conjugate?
3. Find the points of a conic having tangents with slope 1 and the corresponding ideal points on their transform.
4. Find the conic transform for the case $A_5 = A_6 = 0$ directly to avoid the problem with $H = 0$. Is there a parabola \mapsto hyperbola transformation in this case?
5. Using (7.34), construct a table listing the conditions for the A_j when $I > 0$. For example, $I > 0$ and $A_2 > 0 \Rightarrow 1 > r_i$, or $1 < r_i$, $i = 1, 2$, and proceed as for $1 > r_i > 0 \Rightarrow A_2 > A_1 > 0 > A_4$, etc.
6. Find the conic transform of the conjugate hyperbola of that shown in
 - (a) Fig. 7.25.
 - (b) Fig. 7.26.
 - (c) Fig. 7.27.
7. Which conic has as transform the conjugate of the hyperbola shown in the xy plane in Fig. 7.27?
8. What is the image of a hyperbola having a vertical asymptote like the one shown on the left part of Fig. 7.23?
9. Prove that interior points of the conics mapped into hyperbolas are mapped into lines between the two branches. Here the “interior” points of a hyperbola are those between the two branches.
10. Consider hyperbolas having the property that if two lines are completely between the two branches, then the point on the two lines is also between the two branches. Delimit the class of hyperbolas having this property and provide a counterexample of a hyperbola without this property.
11. Obtain the main result in the reverse direction, that is, show that

line-conics in the xy plane \mapsto point-conics in the x_1x_2 plane

and delimit the various cases.

7.6 FT-7 Transforms of Algebraic Curves

Conic transforms are studied for two reasons. One is the ease of use of the Möbius transformations, which is completely general for *quadratics*, the surfaces prescribed by quadratic equations, in *any dimension*. The other is that their detailed analysis serves as a model and guide in the study of the related images of far more complex curves, regions, and their ramifications in the next section.

The class of algebraic curves, those described by polynomial equations, is studied in the field of algebraic geometry ([37], [78], [179], and [86]). An algebraic curve c has many invariants that are properties independent of the particular coordinate system used. Examples are the number of components in its graph, *degree* n , number of *double points* d (points where the curve crosses itself once), *cusps* s , *inflection points* ip , and *bitangents* b (i.e., tangents at *two* points). The results here apply to any curve c and its *dual* curve c^* for *any point* \leftrightarrow line duality. The same symbols with the $*$ superscript denote the dual's invariants. Algebraic curves and their duals have been studied extensively, starting in 1830 by the mathematician and physicist Julius Plücker, who also made other important contributions to the field. His results apply to the class of *Plücker curves* [61, pp. 73–94] having the following properties:

1. c is irreducible and of degree $n \geq 2$, and
2. the singularities of c , \bar{c} are at most simple double points (i.e., points where the curve crosses itself once) and cusps.

Whereas all irreducible quadratics and cubics are *Plücker curves*, there exist quartics, $n = 4$, that are not.

Of interest here are the relations between a curve's invariants and those of its dual. As indicated in the equalities tabulated below, in addition to the *ip* \leftrightarrow *cusp* duality, there is a *bitangent* \leftrightarrow *double point* duality, which we already met in Fig. 7.15. This is reasonable, for the two tangents at a double point map into the two points on a bitangent, which is the double point's image, and in fact, it is true for general and not just algebraic curves. The dual of c is an algebraic curve whose degree n^* depends on n and the invariants d, s as given by the *Plücker class formula*

$$n^* = n(n - 1) - 2d - 3s. \quad (7.40)$$

For $n = 2$ the *Plücker* class formula yields $n^* = 2$ and $s^* = 0$, confirming the conclusions in Section 7.5.

The polynomial describing the dual c^* can be found for *any point* \leftrightarrow *line* duality by two different methods [61, p. 74]. However, with $n^* = O(n^2)$ and the complexity increasing rapidly with the number of nonzero coefficients of the polynomial specifying c , the process is very tedious. All this applies to \bar{c} , which is the image under the particular \parallel -coords duality, when c is an algebraic curve c . As pointed out in Section 7.4, the properties of \bar{c} are immediately available from *Plücker's* results. Together with the qualitative considerations discussed and a good curve plotter, a complete grasp of \bar{c} and its properties can be obtained. This avoids the laborious process of obtaining the polynomial for \bar{c} and its subsequent computation for plotting. There are numerous generalizations of the *Plücker* formulas.

7.7 ♣ FT-8 Convex Sets and Their Relatives

Consider a double cone, as shown in Fig. 7.28, whose base is a bounded convex set rather than a circle. The three type of sections shown are generalizations of the conics and are conveniently called *gconics*. They are one of the following:

a bounded convex set is abbreviated by *bc*,

an unbounded convex set, denoted by *uc*, containing a nonempty set of ideal points whose slope m is in an interval $m \in [m_1, m_2]$, or

a generalized hyperbola, denoted by *gh*, consisting of two full (not segments) lines ℓ_u, ℓ_ℓ , called asymptotes, and two infinite chains: one convex-upward chain c_u above both asymptotes and another convex-downward chain c_ℓ below both asymptotes.

In the past, the corresponding regions were also referred to as *estars*, *pstars*, and *hstars* [99], [110]. Here the *bc* or *uc* are considered as consisting of their interior points and boundaries. A *gh* consists of the points *between* the upper and lower chains and the boundaries. As for the hyperbola, this is best understood in the setting of the projective plane, where the *gh* is a “*bc*” intersecting the ideal line at two points, the asymptotes. In the event that the upper and lower chains have straight-line edges at the endpoints, then these two infinite lines are the asymptotes: ℓ_u joins the leftmost point (i.e., with minimum x) of the upper chain c_u to the

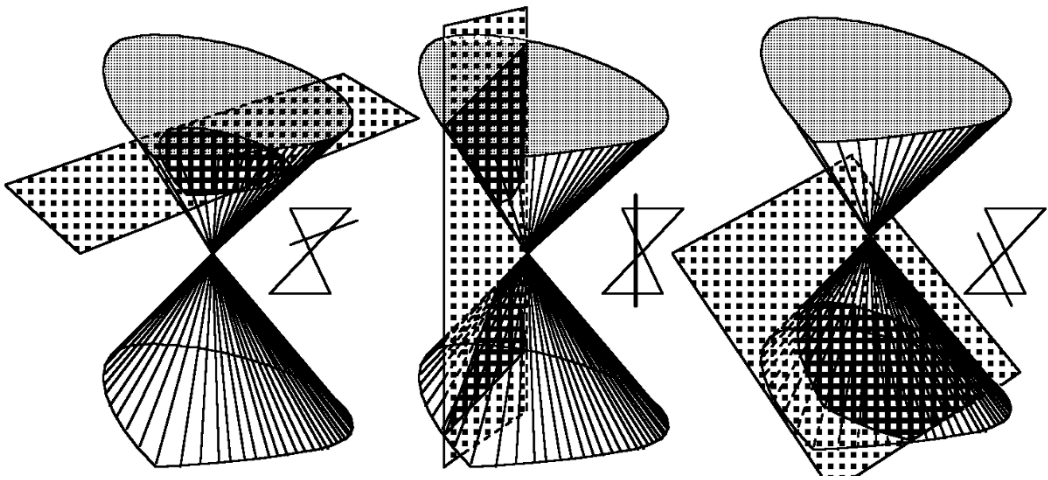


Figure 7.28. Gconics: three types of sections.

Bounded convex *bc* (left), unbounded convex *uc* (right), *gh* regions (two ideal points) (middle).

rightmost point (i.e., with maximum x) of the lower chain c_ℓ and, ℓ_ℓ joins the leftmost point of c_ℓ to the rightmost point of c_u .

It turns out that the conic transforms are special cases of the more general and completely analogous results for *gconics* [99], [110]. In order to study the gconic transforms, it is assumed that their boundaries are point-curves that transform nicely to line-curves, as discussed at the beginning of this chapter. Since the treatment is in \mathbb{P}^2 , the understanding is that convexity is interpreted completely as in the Euclidean \mathbb{R}^2 plane with the line segment (not including the ideal point) joining a pair of points in the set also contained in the convex set.

7.7.1 Gconics and Their Transforms

The insight obtained from the indicator in Section 7.5.2 points the way to the corresponding analysis for the gconics, though there is no explicit relation like (7.34). The discussion pertains to the *supporting lines* of a set's boundary at points where the tangent is not defined. The key convexity property used here is that for any given slope m , a convex set has a unique pair of supporting lines with that slope. As for the conics, finite and infinite line segments are the degenerate cases of a bc and uc respectively, while a pair of intersecting lines is a degenerate gh . These are easily handled directly, so they are excluded from the ensuing treatment. Unless otherwise stated, a gconic is the set consisting of the boundary and its interior.

Theorem 7.7.1. *The images of gconics are gconics.*

1. BC is a $bc \Rightarrow \overline{BC}$ is a gh .

**** Proof.** The boundary $\partial(BC)$ of a bounded convex BC has precisely two points A_u, A_ℓ with supporting lines a_u above a_ℓ having slope $m = 1$. The line w on A_u, A_ℓ divides the boundary into two chains c_u above and c_ℓ below it respectively. The image $\overline{\partial(BC)}$ is split by the two “asymptotes” $\overline{A_\ell}$ and $\overline{A_u}$ into two chains. There exists a point V on $\partial(BC)$ such that $c_u = c_1 \cup c_2$, where $c_1 = A_\ell V$ is a cu chain with $M_1 \geq 1$ and $c_2 = V A_u$ a cd chain with $M_2 < 1$. Hence, by Corollary 7.1.4, $\overline{c_u} = \overline{c_1} \cup \overline{c_2}$ is a cu chain. Similarly, it can be shown that $\overline{c_\ell}$ is a cd chain. Therefore $\overline{\partial(BC)}$ is the boundary of a gh . ■

An example is shown in Fig. 7.29. Further, as seen in Fig. 7.34, details of the mapping of the bc into a gh can be easily tracked. That is, $\overline{A_\ell} \cap \overline{A_u} = \overline{w}$, $\overline{c_1}$ and $\overline{c_2}$ are the parts of $\overline{c_u}$ from the left (which corresponds to $m \geq 1$) and the right of the x axis (for $m > -\infty$) respectively. Similarly for c_ℓ . As for ellipses, interior points of the bc are mapped into lines between the two branches of the gh . This observation has useful consequences discussed at the end of this section.

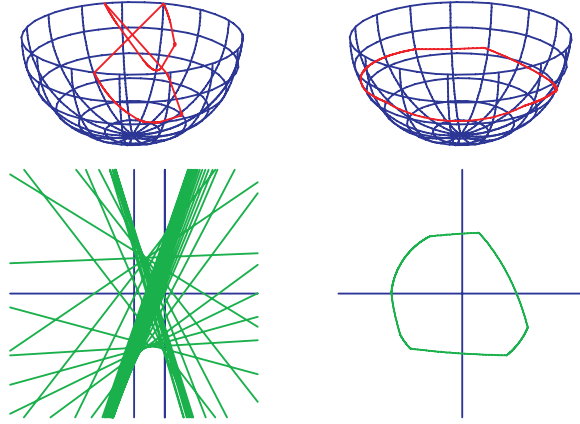


Figure 7.29. A bounded convex set bc always transforms to a gh (generalized hyperbola). This is the generalization of the ellipse \rightarrow hyperbola transform of Fig. 7.22.

2. UC is a $uc \Rightarrow$

- (a) \overline{UC} is a gh if UC does not have an ideal point with slope $m = 1$, i.e., $1 \notin [m_1, m_2]$.
- (b) \overline{UC} is a uc if UC has an ideal point with slope $m = 1$, i.e., $1 \in [m_1, m_2]$.

**** Proof.** (a) UC has one supporting line with slope $m = 1$ at a point $B \in \partial(UC)$. If $m_1 < 0$, $\exists V \in \partial(UC)$ with a vertical supporting line ℓ_v there,³⁵ and if $m_2 < 0$, $\exists H \in \partial(UC)$ with a horizontal supporting line ℓ_H there. Let c_1 be the boundary portions from V upward, c_2 from V to B , c_3 from B to H , and c_4 the remainder from H onward. Note that c_1 is cu , and c_2, c_3, c_4 are cd with $M_1 < 0$, $M_2 \geq 1$, $M_3 \leq 1$, and $M_4 < 1$. Therefore, by Corollary 7.1.4, $\overline{c_1 \cup c_2}$ is a cd above \overline{B} and the vertical line $\overline{P_{m_2}^\infty}$ forming the upper branch of a gh in the xy plane, while $\overline{c_3 \cup c_4}$ is a cu providing the lower branch of the gh below \overline{B} and $\overline{P_{m_1}^\infty}$. The argument still holds if some of the c_i are not empty, that is, when there does not exist B or H . It is easy to prove that the image of an interior point A of UC is a line \overline{A} between the two branches. Hence the strip with endpoints $1/(1 - m_2)$, $1/(1 - m_1)$ contains the vertical lines that are the images of the ideal points in UC ; this strip is also between the upper and lower branches. ■

In the example shown in Fig. 7.30 there is only one ideal point.

³⁵Note that m_1 here corresponds to the *upper* portion of $\partial(UC)$.

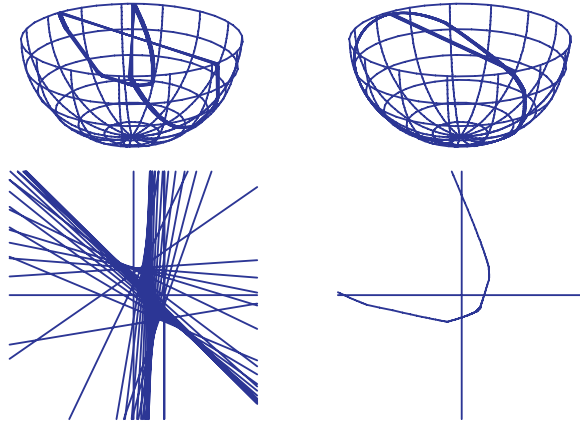


Figure 7.30. An unbounded convex set uc with ideal point slope $m \neq 1$ transforms to a gh . This is the generalization of the conic case in Fig. 7.23.

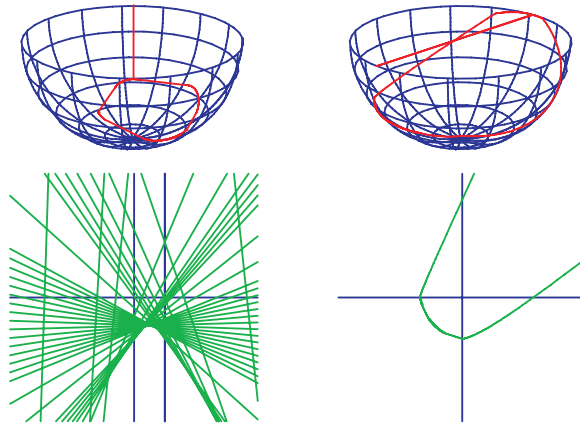


Figure 7.31. A uc with ideal point slope $m = 1$ transforms to a uc ; self-dual case. This is the generalization of the conics case in Fig. 7.24.

**** Proof.** (b) UC has supporting lines with slope $m \in [m_1, m_2]$, where $m_1 < 1 < m_2$. Hence $\exists V \in \partial(UC)$ with a vertical supporting line ℓ_v there. The portion c_1 of $\partial(UC)$ from V downward has $M_1 < 1$ and is cd , while the portion c_2 upward from V has $M_2 \geq 1$ and is cu . Again appealing to the corollary, $\overline{\partial(UC)} = \overline{c_1 \cup c_2}$ is cu . It is formed among other things by the collection of *all* supporting lines having slopes $m \notin [m_1, m_2]$. So as shown in Fig. 7.31, the transform contains at least one ideal point, corresponding to (7.37) for conic parabolas, and is therefore a uc . ■

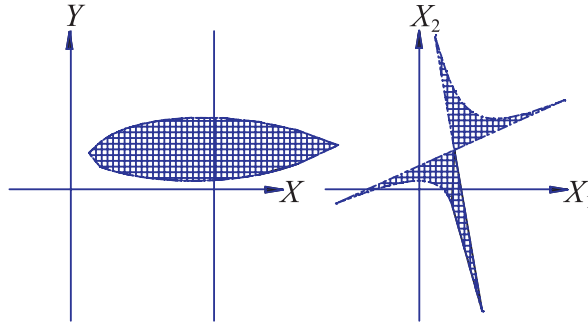


Figure 7.32. A $gh \rightarrow$ to a bc .

Here the asymptotes slopes are $m_1 < 1 < m_2$, with the supporting line's slope $m \in [m_1, m_2]$. This is the generalization of the conic case in Fig. 7.25.

**** Proof.** (a) Since $m_1 < 1 < m_2$, the slope m of the supporting lines of GH is not equal to 1. Hence \overline{GH} is not split and does not have any ideal points \Rightarrow it is a bc whose x ranges in an interval having endpoints $\frac{1}{1-m_1}$ and $\frac{1}{1-m_2}$, the order, of course, being determined by the signs of the m_i . ■

It is worth studying the example shown in Fig. 7.32 below, tracking the images of the various portions of the gh .

**** Proof.** (b) \overline{GH} has at least one ideal point, which is the image of the point of GH where the supporting line has slope $m = 1$, and since GH cannot have any other supporting lines with $m = 1$, it has only one component $\Rightarrow \overline{GH}$ is a uc . ■

**** Proof.** (c) As for the conics, when a *hyperbola* \mapsto *ellipse*, its conjugate *hyperbola* \mapsto *hyperbola*. This case corresponds to a gh “conjugate” to that considered in case (a) as shown in Fig. 7.33. ■

Referring again to Fig. 7.34, consider the bc set shown; let us call it SC , and distinguish it from its subset NSC , where edges AC and DF of SC have been replaced by AB, BC and DE, EF respectively. The gh we see in \parallel -coords is \overline{SC} . It is also the image of the *convex hull* $C(NSC)$ of the nonconvex set NSC . The images of the points B, E and edges AB, BC, ED, EF are lines and points, respectively, and are between the upper and lower branches \overline{SC} . This is the smallest gh that contains \overline{NDS} . In effect, one gets the image of the convex hull of a nonconvex set “for free” in \parallel -coords. The precise statement for this result is the following:

Corollary 7.7.2. *If NSC is a nonconvex set and gh with $\overline{OSC} \supset \overline{NSC}$, then $\overline{OSC} \supseteq C(NSC)$.*

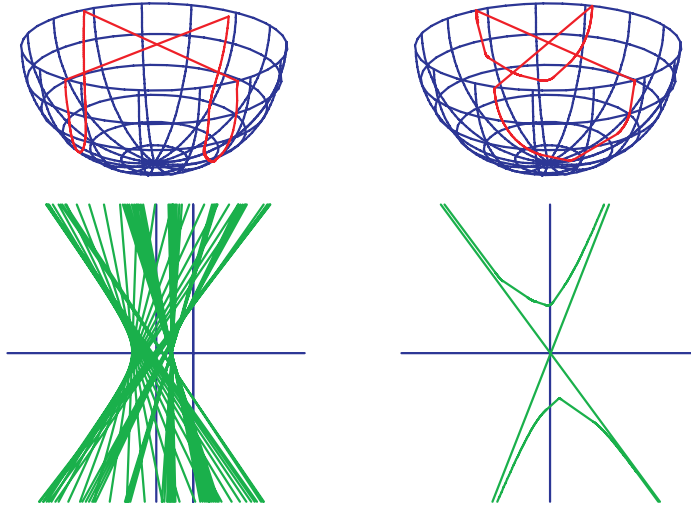


Figure 7.33. ♣ *FT-8e* A $gh \rightarrow gh$; self dual.
Here $1 \notin [m_1, m_2]$, m_i are the asymptotes' slopes. Generalization of Fig. 7.27.

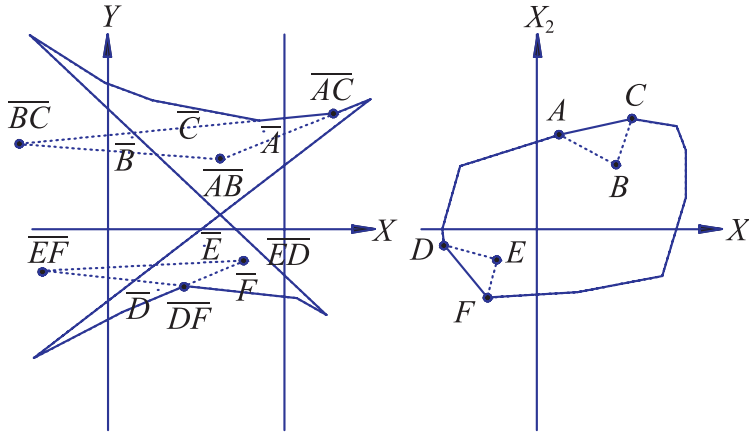


Figure 7.34. Interior points of $bc \rightarrow$ lines between two branches of the corresponding gh .

This result has been used in convex hull construction algorithms [110].³⁶ Here the edges of the bc may be curves, and hence this approach lends itself to the construction of convex hulls where some of the “edges” are curved arches (see [92]).

³⁶The algorithm was improved by Nitzan Giles in 2004 in his course project.

Indications are presented in the next chapter on the feasibility of generalizations to higher dimensions of the result on the appearance of bounded convex sets in $\|\cdot\|$ -coords as generalized hyperbolas.

The generalization of the remarks in Section 7.5.2 as they pertain to ghs is fairly direct. We have seen that the interior points of a bc are mapped into lines that are between the two branches of its image, whereas a gh having vertical asymptotes can only be the image of a uc , as seen in Fig. 7.30. As for the conics, ghs that represent bcs have branches that can be separated by horizontal lines and ghs that represent ghs have branches that can be separated by vertical lines.

The proof of the converse result, \overline{GC} is a gconic $\Rightarrow GC$ is a gconic, is left as an exercise.

Exercises

1. Prove that an unbounded convex set having more than one ideal point has ideal points with direction having slope $m \in [m_1, m_2]$. That is, either m has a single value or it ranges over a complete interval.
2. What is the image of a square centered at the origin with
 - (a) axis-parallel sides,
 - (b) rotated by 45° ?
3. What is the image of the region in the second quadrant above the line $\ell_1 : x_2 = -3/4x_1$ and below the line $\ell_2 : x_2 = -1/4x_1$?
4. What is the image of the region in the fourth quadrant below the line $\ell_1 : x_2 = -3/4x_1$ and above the line $\ell_2 : x_2 = -1/4x_1$?
5. What is the image of the regions in the second and fourth quadrants between the two lines ℓ_1, ℓ_2 ?
6. Study the images of the gconics obtained from a double tetrahedron.
7. Examine the images \overline{GC} of a gconic GC and match the corresponding chains c_i with the corresponding chains $\overline{c_i}$ in \overline{GC} for the case:
 - (a) \overline{GC} is a gh . Distinguish the three cases.
 - (b) \overline{GC} is a uc . How can the image of a uc and a gh be distinguished?
 - (c) \overline{GC} is a bc .
8. Delimit the class of ghs with the property that if two lines are completely between the two branches, then the point on the two lines is also between the two branches. Give a counterexample of a gh without this property.
9. Prove that \overline{GC} is a gconic $\Rightarrow GC$ is a gconic.

7.7.2 ** Operational Dualities with Gconics

While studying the image of a bc as in Fig. 7.34 and experimenting with set operations involving gconics and their images, some interesting De Morgan-like dualities emerged, motivating the following definitions of set operation on two ghs GH_j , $j = 1, 2$, whose two branches have vertical separation and no vertical asymptotes, i.e., they can *only* be images of bcs . Such ghs can be thought of as consisting of complete lines between their two branches.

Definitio 7.7.3. The gh such that $GH \supseteq GH_1 \cup GH_2$ and such that for any other gh , GH' , where $GH' \supseteq GH_1 \cup GH_2 \Rightarrow GH' \supseteq GH$, is called the *outer union*³⁷ of the two ghs . The operation is denoted by $GH_1 \sqcup GH_2$.

An example is shown on the left part of Fig. 7.35. Note that the upper branch of the *outer union* consists of all arches (i.e., lines or curves) belonging to the upper branches of the GH_j on or above all points of the GH_j . Likewise, the lower branch consists of all arches belonging to the lower branches that are on or below all points of the GH_j . Where the endpoints are not ideal, the corresponding asymptote is the line joining the diametrically opposite endpoints of the upper and lower branches. Note that the “innermost” asymptotes of the ghs become the asymptotes of the *outer union*. In general, $GH_1 \sqcup GH_2 \neq GH_1 \cup GH_2$, the *outer union* being the smallest gh containing the union of the two ghs .

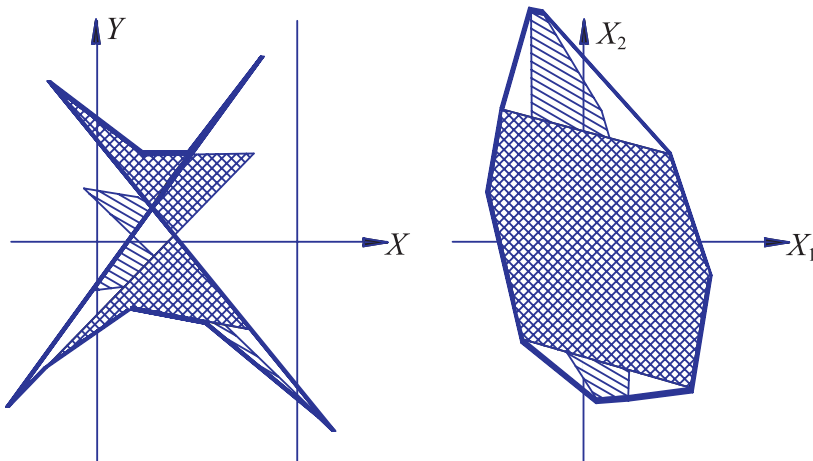


Figure 7.35. The convex union (also called “convex merge”) of bcs . It corresponds to the outer union of their images (ghs).

³⁷This differs from the nomenclature in [110].

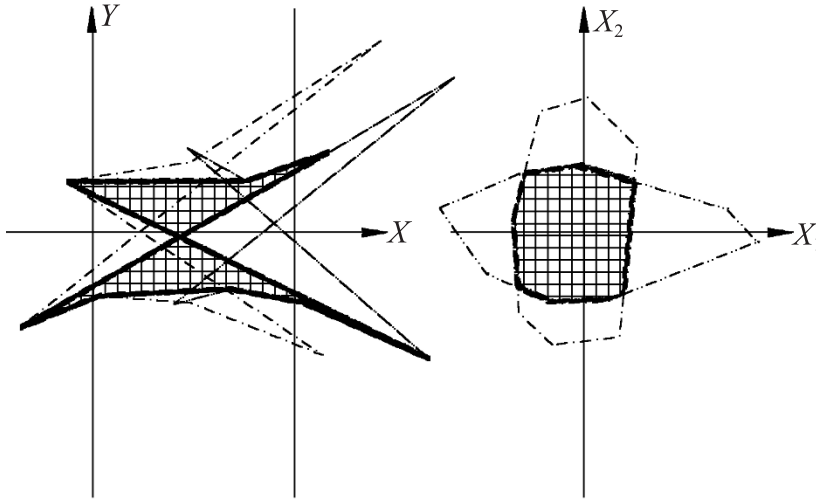


Figure 7.36. Inner intersection and intersections are dual.

Definitio 7.7.4. The *gh* such that $GH \subseteq GH_1 \cap GH_2$ and such that for any other *gh*, GH' , where $GH' \subseteq GH_1 \cap GH_2 \Rightarrow GH' \subseteq GH$, is called the *inner intersection* of the two *ghs*. The operation is denoted by $GH_1 \sqcap GH_2$.

An example is shown on the left part of Fig. 7.36. The *inner intersection* is the *gh* whose upper branch is the *cd* curve consisting of all vertices of the *ghs* and the arches (i.e., lines or curves) joining them on either of the upper branches. Likewise, the lower branch is the *cu* curve consisting of all vertices and arches joining them that are above all points of the lower branches. The asymptotes are the ideal points, if any. Otherwise, they are the lines joining the diametrically opposite endpoints of the upper and lower chains. Here the asymptotes may not belong to any of the *ghs*. In general, $GH_1 \sqcap GH_2 \neq GH_1 \cap GH_2$. It is clear from the definition that the *inner intersection* is the largest *gh* contained in the intersection of the two *ghs*. Further, whereas all vertices of the *inner intersection* are vertices of the GH_j , some of the edges joining such vertices are not edges of the GH_j . Roughly speaking, the *outer union* of two *ghs* yields a *gh* whose branches contain the *outermost* portions of the two *ghs* and any additional edges needed to form the *gh*, whose significance will be shortly understood. This is also the reason why it does not always coincide with the *set union* of the *ghs*. The *inner intersection*, by contrast, contains the *innermost* portions of the two *ghs*. The results below, stated in terms of two *bcs* BC_i , $i = 1, 2$, where $\mathcal{C}(S)$ denotes the *convex hull* of the set S , clarify the motivation for these definitions.

Theorem 7.7.5. $\overline{\mathcal{C}(BC_1 \cup BC_2)} = \overline{BC_1} \sqcup \overline{BC_2}$.³⁸

Proof. The proof is immediate. The convex merge $\mathcal{C}(BC_1 \cup BC_2)$ is the smallest convex set containing $BC_1 \cup BC_2 \Rightarrow \overline{\mathcal{C}(BC_1 \cup BC_2)}$ is the smallest *gh* containing $\overline{BC_1} \cup \overline{BC_2}$, whence the result. ■

Notice in Fig. 7.35 that all lines in $\partial(\overline{BC_1} \sqcup \overline{BC_2})$ represent vertices of the *ghs*, whereas any additional vertices represent the additional edges needed to form the *convex merge*. As expected, the asymptotes represent the vertices of $\mathcal{C}(BC_1 \cup BC_2)$ where there are supporting lines with slope $m = 1$.

Theorem 7.7.6. $\overline{BC_1 \cap BC_2} = \overline{BC_1} \cap \overline{BC_2}$.

Proof. By definition, the *inner intersection* is the largest *gh* contained in $\overline{BC_1} \cup \overline{BC_2}$. Hence $\overline{BC_1 \cap BC_2}$ is the largest convex set contained in the inverse image of $\overline{BC_1} \cap \overline{BC_2}$, which is also convex. So the two convex sets must be equal and therefore their images are equal. ■

As seen in Fig. 7.36, the edges of $\overline{BC_1 \cap BC_2}$ that are not edges of any of the *ghs* represent the new vertices formed by the intersection of the corresponding *bcs*. Note that the asymptotes here represent the two new vertices of the *convex merge* and where the supporting lines have slope $m = 1$. An easy induction proves the following theorem.

Theorem 7.7.7. $\overline{\mathcal{C}(\cup_{i=1}^n BC_i)} = \sqcup_{i=1}^n \overline{BC_i}$.

Since the union of a countable collection of *bcs* may not be a *bc*, this theorem is in general true for finite collections. The corresponding result for *outer unions* is true for countable collections of *bcs*.

Theorem 7.7.8. $\overline{\cap_{i=1}^\infty BC_i} = \cap_{i=1}^\infty \overline{BC_i}$.

These interesting operational dualities have some algorithmic applications [110].

Exercises

1. Give an example where the *outer union* of two *ghs* is not the union of the *ghs*.
2. Give an example where the *inner intersection* is not the intersection of the *ghs*.
3. Find the image of the intersection of the two ellipses $\frac{x_1}{a^2} + \frac{x_2}{b^2} = 1$ and $\frac{x_1}{b^2} + \frac{x_2}{a^2} = 1$.
4. Find the image of the convex merge of the two ellipses above.

³⁸This operation is also called the *convex merge* of the two sets BC_i .

5. Find the image of the convex merge of two disjoint squares. Investigate various options concerning position and orientation and the effects on the images.
6. Prove
 - (a) Theorem 7.7.7.
 - (b) Theorem 7.7.8.
7. Consider the general third-degree curve $F(x_1, x_2) = \sum_{i+j=0}^3 a_{ij}x_1^i x_2^j$ and
 - (a) show that it can be written in matrix form as

$$F(x_1, x_2) = (x_2^2, x_1x_2, x_1^2, x_2, x_1, 1)\mathbf{A} \begin{pmatrix} x_2^2 \\ x_1x_2 \\ x_1^2 \\ x_2 \\ x_1 \\ 1 \end{pmatrix}, \quad (7.41)$$

where \mathbf{A} is a symmetric 6×6 nonsingular matrix.

- (b) Deduce the degree of the image curve from this equation.
- (c) (*Hard*) Apply the transformations of (7.7) to (7.41) to obtain the image curve.

This page intentionally left blank

Proximity of Lines, Planes, and Flats

8.1 **FT-1** Motivation and a Topology for Proximity

In order to apply the results of the representation of flats by indexed points, their behavior in the presence of errors needs to be understood. This was briefly touched on in Chapter 5, where we saw that small variations in the coefficients of a plane's equation yield a pair of closely spaced point clusters representing the corresponding “close” planes. The encouraging message is that “closeness” is visually evident. To pursue this in depth, we need some concepts discussed in Chapter 7. For this reason, we did not continue with this treatment earlier. While there are many sources of errors in the applications, from our viewpoint it suffices to consider the accumulated errors in terms of the resulting variations $c_i \in [c_i^-, c_i^+]$ in the coefficients c_i of the plane's equations. This in turn generates a whole *family* \mathcal{F} of “close” flats. Even in R^3 , the direct visualization of such a family of planes is difficult. The challenge is not only the visualization but also the construction of tools for appraising the proximity, referred to as “error tolerancing,” of the flats in terms of the errors' ranges $|c_i^- - c_i^+|$.

Let us repeat the earlier experiment more systematically and examine the family of “close” planes

$$\{\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0, c_i \in [c_i^-, c_i^+], c_i^- < c_i^+\}.$$

Computing the two-point representation in $\|\cdot\|$ -coords of some of these planes, we see in Fig. 10.43 the corresponding pair of point clusters. Closeness is apparent, and more significantly, the distribution of the points is not “chaotic.” The outline of two polygonal patterns can be discerned. Not only is the family of planes “visualizable,”



Figure 8.1. Pair of point clusters representing close planes. Note the hexagonal patterns.

but also the variations in several directions. Might it be possible to see, estimate, and compare errors?

The proper setting for treating the proximity of objects in a family \mathcal{F} is by means of a *topology* [138]. Every object $f \in \mathcal{F}$ is endowed with a collection of subsets $\mathcal{N}_f = \{N | N \subset \mathcal{F}\}$, called *neighborhoods*, such that

1. each $f \in \mathcal{F}$ has an \mathcal{N}_f ,
2. $f \in N \forall N \in \mathcal{N}_f$,
3. if $N_1 \in \mathcal{N}_f$ and $N \supset N_1$, then $N \in \mathcal{N}_f$,
4. if $N_1, N_2 \in \mathcal{N}_f$, then $N_1 \cap N_2 \in \mathcal{N}_f$.

Then proximity is gauged with respect to the topology in the sense that f_1 and f_2 are “close” if $\exists N \in \mathcal{N}_{f_1}$ such that $f_2 \in N$ and hence also $N \in \mathcal{N}_{f_2}$. Intuitively, we would like a neighborhood of a flat f to be like a “template,” a region covered by all flats close to f . In dealing with such regions, the results on *gconics* from Chapter 7 turn to be helpful, and for this reason this subject is not included, in Chapter 5 on planes.

8.2 Proximity of Lines and Line Neighborhoods

We start the development by looking into families of lines. Is it reasonable to consider lines in \mathbb{R}^N being “close” if the coefficients of their equations are close?³⁹ In \mathbb{R}^2 consider the collection of lines

³⁹This differs from the proximity studied in Chapter 4 in terms of the minimum distance between pairs of lines.

$$\mathcal{F} = \{\ell | \ell : c_1 x_1 + c_2 x_2 = 1, c_1, c_2 \in \mathbb{R}\}.$$

The value of the constant c_0 on the right-hand side here is taken as 1 but is allowed to vary later. Of interest is the collection of subsets $\mathcal{N}_\ell = \{NL | NL \subset \mathcal{F}\}$, where

$$NL = \{\ell | \ell : c_1 x_1 + c_2 x_2 = 1, \quad c_i \in [c_i^-, c_i^+] \quad i = 1, 2\}. \quad (8.1)$$

It is convenient to denote the endpoints of the interval $[c_i^-, c_i^+]$ by $sign_i$, the sign of the superscript of c_i^\pm ; that is, $sign_i = -$ and $sign_i = +$ stand for the lower and upper bounds respectively of c_i . The extreme lines obtained by the four different combinations of $(sign_1, sign_2)$ are

$$\left. \begin{aligned} (-, -) : c_1^- x_1 + c_2^- x_2 &= 1, \\ (-, +) : c_1^- x_1 + c_2^+ x_2 &= 1, \\ (+, -) : c_1^+ x_1 + c_2^- x_2 &= 1, \\ (+, +) : c_1^+ x_1 + c_2^+ x_2 &= 1. \end{aligned} \right\} \quad (8.2)$$

These lines bound the region $R_{NL} \subset \mathbb{R}^2$ covered by all the lines of NL , a “template” referred to earlier. An example is shown on the left part of Fig. 8.2, where the extreme lines in (8.2) are constructed from the points

$$P_1^+ = \left(\frac{1}{c_1^+}, 0 \right), \quad P_1^- = \left(\frac{1}{c_1^-}, 0 \right), \quad P_2^+ = \left(0, \frac{1}{c_2^+} \right), \quad P_2^- = \left(0, \frac{1}{c_2^-} \right).$$

The region R_{NL} resembles a hyperbola and is in fact a generalized hyperbola, a *gh* in the terminology of Chapter 7. All lines in NL are enclosed between two opposite convex unbounded regions with two “asymptotes,” the lines $(+, -)$, $(-, +)$, joining them at two ideal points.

Clearly, as $|c_i^- - c_i^+| \rightarrow 0$ the corresponding lines become more coincident, and it is fair to consider them as “close.” It is legitimate, then, to consider the R_{NL} as candidates for “line-neighborhoods” of close lines. Yet no matter how small the $|c_i^- - c_i^+|$ are, the region R_{NL} is unbounded and becomes wider as the x_i increase. Perhaps it is more manageable to study the situation in the *parametric “dual” space* with the point $(c_1, c_2) \rightarrow \ell : c_1 x_1 + c_2 x_2 = 1$ line correspondence. Then the rectangles $[c_1^-, c_1^+] \times [c_2^-, c_2^+] \subset \mathbb{R}^2$ are alternative candidates for “line-neighborhoods.” Pursuing the idea further, in some applications, *points* are given, together with some lines. The task includes the construction of the “best” (according to some criterion) line from these points. Skipping the details, the important observation here is that we need to work with and view *both* points and lines. This

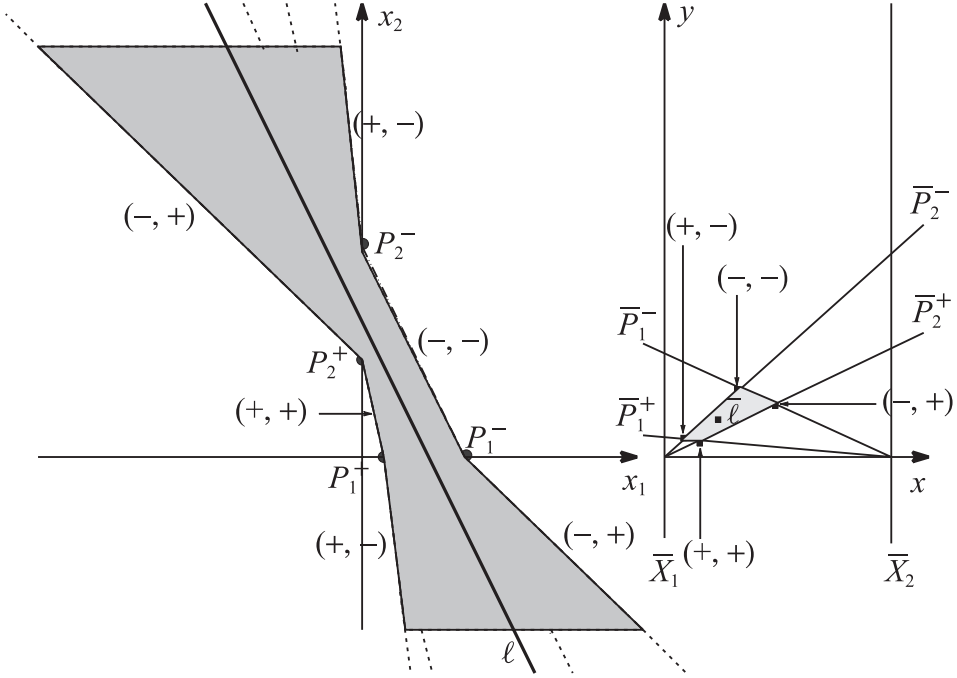


Figure 8.2. A Line neighborhood.

On the left is a region covered by lines “close” to ℓ , and on the right are the points in $\|\text{-coords}$ representing the lines in this region. With reference to the nomenclature and results in Chapter 7, this is an instance of a *gh* (a “generalized hyperbola”) on the left that is mapped into a *bc*, here a bounded convex quadrilateral (right)

can be accomplished using $R_N \subset \mathbb{R}^2$ but is not possible using the parameter space in which *only* the lines, represented by points, are available. Checking the situation in $\|\text{-coords}$, shown on the right part of Fig. 8.2, the unbounded region R_{NL} is transformed into a simple *convex* quadrilateral; let us call it \overline{NL} . From the *point* \leftrightarrow *line* duality, *both* points and lines are available to work with.

The intention is to form a topology; therefore the possible intersections between different quadrilaterals \overline{NL} need to be examined. Let

$$NL_j = \{\ell | \ell : c_{1,j}x_1 + c_{2,j}x_2 = 1, \quad c_{i,j} \in [c_{i,j}^-, c_{i,j}^+] \quad i = 1, 2, j = 1, 2\}, \quad (8.3)$$

and the intersection operation $N = N_1 \cap N_2$ is defined by

$$NL = \{\ell | \ell : c_1x_1 + c_2x_2 = 1, \quad c_i \in [c_i^-, c_i^+] = [c_{i,1}^-, c_{i,1}^+] \cap [c_{i,2}^-, c_{i,2}^+] \quad i = 1, 2\}. \quad (8.4)$$

The different intersection outcomes are these:

1. if $[c_i^-, c_i^+] = \emptyset$ for any i , then $NL = \emptyset$,
2. if $[c_i^-, c_i^+] = c_i$ for $i = 1, 2$, then $NL = \{\ell\}$ for $\ell : c_1x_1 + c_2x_2 = 1$;
3. if $[c_1^-, c_1^+] = c_1$ and $c_2^- < c_2^+$, then NL is a “fan” of lines on the point $P_1 = (\frac{1}{c_1}, 0) \Rightarrow \overline{NL}$ is a segment on \bar{P}_1 . Similarly, if $[c_2^-, c_2^+] = c_2$ and $c_1^- < c_1^+$, \overline{NL} is a segment on \bar{P}_2 where $P_2 = (\frac{1}{c_2}, 0)$; or
4. \overline{NL} is a quadrilateral contained in both \overline{NL}_1 and \overline{NL}_2 .

Though the set intersection of two quadrilaterals may be a triangle, the intersection operation defined by (8.4) excludes that possibility (see the exercises below). With each of these cases a legitimate \overline{NL} , it is clear that the family $\mathcal{NL}_\ell = \{\overline{NL} | NL \subset \mathcal{F}\}$ is a topology of line-neighborhoods.

By contrast, the intersection of two R_{NL} is not in general such a region. Four distinct R_{NL} are shown in Fig. 8.3, and though some pairs have a nonempty intersection, it does not contain any complete lines. This is clearly seen from the corresponding \overline{NL} , which are disjoint. So the regions R_{NL} considered as *point sets* cannot serve as neighborhoods for a topology for lines. The alternative is to consider

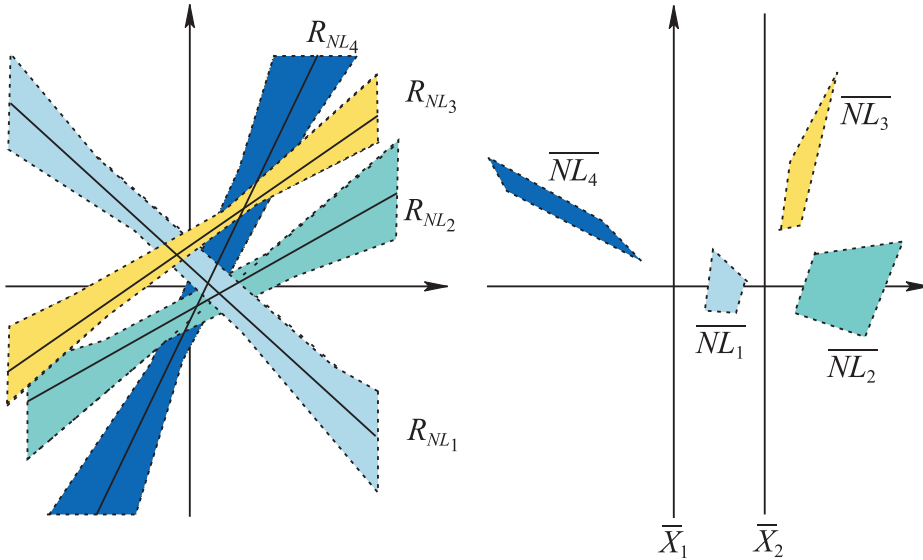


Figure 8.3. Several line neighborhoods.

The regions R_{NL_i} , $i = 1, 2, 3, 4$, covered by four families of lines in orthogonal coordinates and their images \overline{NL}_i in $||$ coordinates.

R_{NL} as consisting of the *complete lines* covering it, greatly complicating matters. Instead, we choose to work with their images \overline{NL} , which satisfy all the requirements for a useful topology of lines. Additionally, with $||$ -coords we are not limited by the dimensionality of the space either for lines or more general flats. This leads to some surprising, elegant, and very useful generalizations.

Before getting to that, we enlarge the class of lines in N replacing the 1 by c_0 and allowing it to vary within an interval

$$NL = \{\ell | \ell : c_1 x_1 + c_2 x_2 = c_0, \quad c_i \in [c_i^-, c_i^+] \quad i = 0, 1, 2\}. \quad (8.5)$$

Each such line ℓ is represented by the point

$$\bar{\ell} = \left(\frac{c_2}{c_1 + c_2}, \frac{c_0}{c_1 + c_2} \right). \quad (8.6)$$

In order to study the resulting neighborhood \overline{NL} , first we take $c_0 = c_0^-$. This is basically the case already studied, and the corresponding neighborhood, whose construction is shown in the lower part of Fig. 8.4, is formed using the points with values of c_0^- in the numerator. The vertices of the quadrilateral are labeled by $(\text{sign}_1, \text{sign}_2, -)$, since they are the points representing the four extreme lines as before. Next the value $c_0 = c_0^+$ is chosen and the corresponding quadrilateral is constructed now using the points with values of c_0^+ in the numerator with vertices $(\text{sign}_1, \text{sign}_2, +)$. Inspecting the y -coordinate of the point given by (8.6), it is clear that for the intermediate values $c_0 = [c_0^-, c_0^+]$ between the two endpoints,

$$\frac{c_0^-}{c_1 + c_2} \leq \frac{c_0}{c_1 + c_2} \leq \frac{c_0^+}{c_1 + c_2},$$

the vertices of the quadrilaterals are translated upward from the quadrilateral obtained from c_0^- until the vertices for c_0^+ with the same x -coordinate. The resulting \overline{NL} is now a *hexagon* with the leftmost and rightmost edges vertical. The vertices going counterclockwise are $(-, +, -)$, $(+, +, -)$, $(+, -, -)$, $(+, -, +)$, $(-, -, +)$, $(-, +, +)$. For clarity the three quadrilaterals corresponding to c_0^- , c_0^+ , and $c_0 \in [c_0^-, c_0^+]$ are also shown separately in the upper portion of Fig. 8.5 with fewer details.

We reexamine the prospects for a topology using the hexagons as line-neighborhoods, starting with

$$NL_{j,0} = \{\ell | \ell : c_{1,j} x_1 + c_{2,j} x_2 = c_0, \quad c_{i,j} \in [c_{i,j}^-, c_{i,j}^+], \quad i = 0, 1, 2, j = 1, 2\}, \quad (8.7)$$

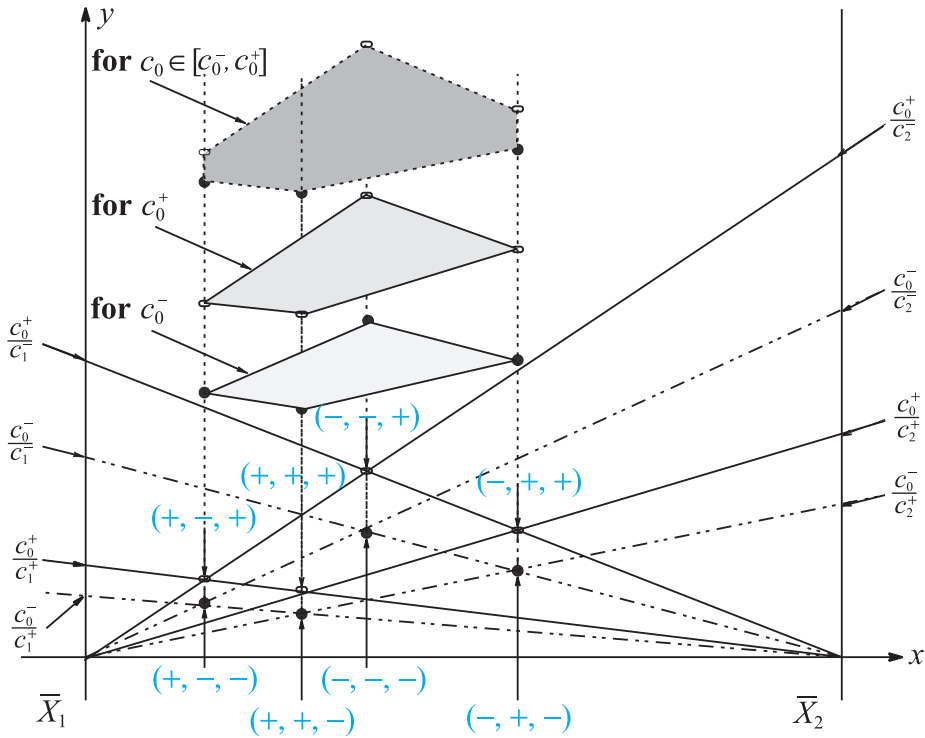


Figure 8.4. Construction of a neighborhood \overline{NL} for $NL = \{\ell \mid c_1x_1 + c_2x_2 = c_0, c_i \in [c_i^-, c_i^+], i = 0, 1, 2\}$. The “exploded” view also shows the quadrilaterals \overline{NL}_- for $c_0 = c_0^-$ and \overline{NL}_+ and $c_0 = c_0^+$, whose vertices are marked with black and blank oval vertices respectively. The complete \overline{NL} is a *hexagon* having two vertical edges, three vertices from \overline{NL}_- , the lowest ones, and three vertices from \overline{NL}_+ .

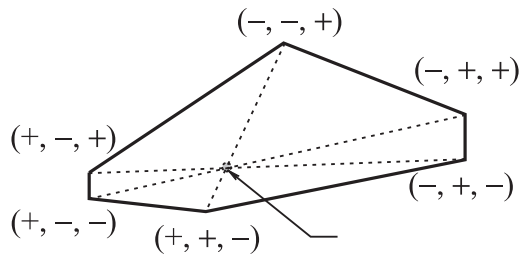


Figure 8.5. The hexagon of the previous figure without the construction details. The arrow shows that the three diagonals intersect at a point (Exercise 6).

for the intersection $NL = NL_1 \cap NL_2$ defined by

$$\begin{aligned} NL_0 &= \{\ell | \ell : c_1 x_1 + c_2 x_2 = c_0, \ c_i \in [c_i^-, c_i^+]\} \\ &= [c_{i,1}^-, c_{i,1}^+] \cap [c_{i,2}^-, c_{i,2}^+], \ i = 0, 1, 2. \end{aligned} \quad (8.8)$$

♣ FT-1e

The different outcomes of the intersection are then:

1. If $[c_i^-, c_i^+] = \emptyset$ for any i , then $NL = \emptyset$.
2. If $[c_0^-, c_0^+] = 0$ and $[c_i^-, c_i^+] = 0$ for either $i = 1$ or $i = 2$, then ℓ is on the x_2 -axis if $i = 1$ and on the x_1 -axis if $i = 2$.
3. If $[c_0^-, c_0^+] = 0$ and $[c_i^-, c_i^+] \neq 0$ for both $i = 1, 2$, then R_{NL} consists of two overlapping “fans” of lines at the origin $(0, 0)$, so \overline{NL} consists of two overlapping intervals on the x -axis. In effect, this is the case where the hexagon (see Fig. 8.5) collapses to a horizontal line.
4. If $[c_0^-, c_0^+] = c_0 \neq 0$, then this yields a quadrilateral \overline{NL} with the cases enumerated earlier.
5. If $c_0^- < c_0^+$ and c_i is constant for either $i = 1, 2$, then \overline{NL} is a quadrilateral with two vertical sides and R_{NL} consists of two strips each having constant width.
6. \overline{NL} is a hexagon contained in both $\overline{NL}_{1,0}$ and $\overline{NL}_{2,0}$.

Again we can conclude that the family of hexagons \overline{NL}_0 constitutes a topology of line-neighborhoods, since all the enumerated situations are also specific instances of $NL_{j,0}$. Allowing the variation of $c_0 \in [c_0^-, c_0^+]$ is the most general case for a family of lines. To complete the study there still remains to examine the situation in which the sum $c_1 + c_2 = 0$ with some of the corresponding lines, having slope 1, is represented by ideal points.

The generalization to \mathbb{R}^N is direct, a family of close lines is represented by $N - 1$ regions, and it comes out as a special case in what follows (see also Exercise 8 below). For earlier treatments of line-neighborhoods and topologies for flats in $||$ -coords see [107] and [23]. The exposition here is adapted from [131].

Exercises

1. Show that the intersection, as defined by (8.4), of two quadrilaterals \overline{NL}_j cannot be a triangle. Draw the R_{NL} corresponding to a triangular \overline{NL} contained between the \bar{X}_1 and \bar{X}_2 axes and describe the family of lines NL .
2. Draw examples of the quadrilaterals \overline{NL}_j for $j = 1, 2$ and the R_{NL_j} for all the different cases of the intersections $\overline{NL} = \overline{NL}_1 \cap \overline{NL}_2$.
3. Show that when \overline{NL} , as given by (8.1), is a quadrilateral, it is also convex.

4. Construct an example in which \overline{NL} is not convex and describe the region and that of R_{NL} .
5. Show that when \overline{NL}_0 , as given by (8.7), is a hexagon, it is also convex.
6. Show that the three diagonals of the hexagon having exactly opposite signs (i.e., $(-, +, +)$ is joined to $(+, -, -)$) intersect at the same point; see also Fig. 8.5.
7. Draw examples of the hexagons $\overline{NL}_{j,0}$, for $j = 1, 2$ and the $R_{NL_{j,0}}$ for all the different cases of the intersections $\overline{NL}_0 = \overline{NL}_{1,0} \cap \overline{NL}_{2,0}$.
8. Generalize to \mathbb{R}^N the representation of a family of close lines and provide an algorithm for constructing the line neighborhoods; this is a good topic for a *course project*.

8.3 **FT-2** Proximity of Hyperplanes

8.3.1 Formulation of the Problem in \mathbb{R}^N

In \mathbb{R}^3 , for the plane

$$\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0, \quad (8.9)$$

$$\begin{aligned} \bar{\pi}_{1'} = \bar{\pi}_{1'23} &= \left(\frac{c_2 + 2c_3 + 3c_1}{c_1 + c_2 + c_3}, \frac{c_0}{c_1 + c_2 + c_3} \right) \\ &= (1, 0) + \left(\frac{c_3 + 2c_1}{c_1 + c_2 + c_3}, \frac{c_0}{c_1 + c_2 + c_3} \right), \end{aligned} \quad (8.10)$$

an observation showing that $\bar{\pi}_{1'} = \bar{\pi}_{1'23}$ can be obtained from

$$\bar{\pi}_{0'} = \bar{\pi}_{123} = \left(\frac{c_2 + 2c_3}{c_1 + c_2 + c_3}, \frac{c_0}{c_1 + c_2 + c_3} \right)$$

by a shift and the counterclockwise cyclic permutation $c_1 \rightarrow c_3, c_3 \rightarrow c_2, c_2 \rightarrow c_1$. This combinatorial relation allows for a reduction in the number of indexed points that need to be considered. As a reminder, *any* two points in the xy plane labeled $\bar{\pi}_{123}, \bar{\pi}_{1'23}$ with the same y -coordinate are required for a valid representation of π , the horizontal distance between them being proportional to c_1 . Tracing the permutation between the c_k also clarifies that the *value* of $\bar{\pi}_{1'23}$ can *not* be obtained from $\bar{\pi}_{123}$ without additionally specifying a value for c_1 . Continuing,

$$\bar{\pi}_{2'} = \bar{\pi}_{1'2'3} = (2, 0) + \left(\frac{c_1 + 2c_2}{c_1 + c_2 + c_3}, \frac{c_0}{c_1 + c_2 + c_3} \right), \quad (8.11)$$

as obtained from $\bar{\pi}_{0'} = \bar{\pi}_{123}$ via a shift and the clockwise cyclic permutation $c_1 \rightarrow c_2, c_2 \rightarrow c_3, c_3 \rightarrow c_1$. Finally,

$$\bar{\pi}_{3'} = \bar{\pi}_{1'2'3'} = (3, 0) + \left(\frac{c_2 + 2c_3}{c_1 + c_2 + c_3}, \frac{c_0}{c_1 + c_2 + c_3} \right) = (3, 0) + \bar{\pi}_{0'}. \quad (8.12)$$

The analogous relations hold for \mathbb{R}^N . For a hyperplane $\pi : c_1x_1 + c_2x_2 + \cdots + c_Nx_N = c_0$ let $\mathbf{c} = (c_1, c_2, \dots, c_N)$ be the N -tuple of the coefficients and for an N -tuple $Z = (z_1, z_2, \dots, z_N)$, denote the sum by $S(Z) = \sum_{j=1}^N z_j$. According to Corollary 5.4.2, the first and second points representing π are

$$\bar{\pi}_{12\dots N} = \bar{\pi}_{0'} = \left(\frac{c_2 + \cdots + (k-1)c_k + \cdots + (N-1)c_N}{S}, \frac{c_0}{S} \right), \quad (8.13)$$

$$\bar{\pi}_{1'} = \left(\frac{Nc_1 + c_2 + \cdots + (k-1)c_k + \cdots + (N-1)c_N}{S}, \frac{c_0}{S} \right), \quad (8.14)$$

where $S = S(\mathbf{c}) = \sum_{j=1}^N c_j$. Here

$$\begin{aligned} \bar{\pi}_{1'} &= (1, 0) \\ &+ \left(\frac{c_3 + 2c_4 + \cdots + (k-2)c_{k-1} + \cdots + (N-2)c_N + (N-1)c_1}{S}, \frac{c_0}{S} \right) \end{aligned} \quad (8.15)$$

is obtained from $\bar{\pi}_{0'}$ by a shift and the counterclockwise cyclic permutation $c_1 \rightarrow c_N \rightarrow c_{N-1} \rightarrow \cdots \rightarrow c_k \rightarrow c_{k-1} \rightarrow \cdots \rightarrow c_2 \rightarrow c_1$ generalizing (8.10), a process that can be continued inductively for the remaining

$$\begin{aligned} \bar{\pi}_{i'} &= \left(\frac{Nc_1 + (N+1)c_2 + \cdots + (N+i-1)c_i + ic_{(i+1)} + \cdots + (N-1)c_N}{S}, \frac{c_0}{S} \right), \end{aligned} \quad (8.16)$$

obtained from $\bar{\pi}_{0'}$ by a shift of $(i, 0)$ and an appropriate permutation of the c_k . This is applied, amplified, and better understood in Section 8.3.6.

We proceed now with the study of the class of all hyperplanes

$$\mathcal{F} = \{\pi | \pi : c_1x_1 + c_2x_2 + \cdots + c_Nx_N = 1, c_i \in R \text{ for } i = 1, 2, \dots, N\},$$

and the collection of neighborhoods $\mathcal{N}_\pi = \{NH | NH \subset \mathcal{F}\}$, where

$$NH = \{\pi | \pi : c_1x_1 + c_2x_2 + \cdots + c_Nx_N = 1, \quad c_i \in [c_i^-, c_i^+] \quad i = 1, 2, \dots, N\}. \quad (8.17)$$

Note that for convenience, the value of $c_0 = 1$ is used and is adjusted later to arbitrary values as for line neighborhoods. In the study of the $N - 1$ regions composing \overline{NH} , in view of the previous reduction, it suffices to investigate the region $\Omega = \overline{NH}_0$ containing the points $\bar{\pi}_0$. These are the function values of

$$f_N : \mathbb{R}^N \mapsto \mathbb{R}^2$$

in x, y coordinates,

$$f_N(\mathbf{c}) = (x(\mathbf{c}), y(\mathbf{c})) = \left(\frac{\sum_{j=1}^N (j-1)c_j}{S}, \frac{1}{S} \right), \quad (8.18)$$

reverting to homogeneous coordinates to accommodate ideal points. The subscript “0” is dropped when the context is clear. The c_j range over the N -dimensional “box” is

$$\mathbf{B} = [c_1^-, c_1^+] \times \cdots \times [c_N^-, c_N^+] \subset \mathbb{R}^N.$$

The stage is now set for the task ahead: to understand the properties of f_N , its action on B ,

$$\mathbf{B} \xrightarrow{f_N} \Omega = f_N(\mathbf{B}) \subset \mathbb{P}^2,$$

and the structure of Ω representing the family of hyperplanes in \mathcal{N} .

The previously studied line neighborhoods serve as a guide and prototype. The region of “close” lines, Fig. 8.2, is a *gh* (generalized hyperbola), and from the previous chapter we know that its \parallel -coords image can be any *bc* (bounded convex set), the case covered so far, *uc* (unbounded convex set), or *gh*. The region in \mathbb{R}^N covered by “close” hyperplanes is a complex N -dimensional *gh* whose image in \parallel -coords consists of $N - 1$ regions in \mathbb{R}^2 . As will be seen, these are also *bc*, *uc*, *gh*. From these regions the properties of “close” hyperplanes can be ascertained and visualized without loss of information.

8.3.2 Outline of the Development

The key steps in the ensuing are outlined here so that the development will not be obscured by the details.

1. At first, only the region Ω containing the points $\bar{\pi}_{0'}$ need be considered with the constant on the right-hand side $c_0 = 1$.
2. Analysis of the function f_N in Section 8.3.3 enables the construction of Ω as the intersection of N “fan” regions shown in Fig. 8.10. The result is a polygonal region whose vertices are the values of f_N at $2N$ key vertices of the hyperbox \mathbf{B} given by (8.45) and traversed along the path (8.46).
3. The polygonal regions are either convex bounded or unbounded or gh with $2N$ vertices depending on the intersection of \mathbf{B} with the hyperplane $\pi_c : c_1 + c_2 + \cdots + c_N = 0$ shown in Fig. 8.24. The computational complexity of the construction is $O(N)$.
4. The remaining $N - 2$ regions are then easily obtained by variations in the traversal of the path in (8.46).
5. Allowing $c_0 \neq 1$ is dealt with following (8.55), adding two vertical edges to the previous polygonal region.
6. In the last section, points within the regions are matched, so that together they represent a valid element (i.e., plane or hyperplane) of the family. This serves as a prototype for the equivalent problem in the surface representation.

 **FT-2e**

8.3.3 Properties of f_N

The coordinate x of f_N can be found in terms of the second coordinate $y = 1/S$ via the decomposition below, for $k = 1, 2, \dots, N$, as

$$\begin{aligned}
 x &= \frac{\sum_{j=1}^N (j-1)c_j}{S} = \frac{(k-1)S}{S} + \frac{(1-k)S + \sum_{j=1}^N (j-1)c_j}{S} \\
 &= (k-1) + \frac{S - \sum_{j=1}^N kc_j - S + \sum_{j=1}^N jc_j}{S} = (k-1) + \frac{\sum_{j=1}^N (j-k)c_j}{S},
 \end{aligned}$$

yielding the linear relation

$$\bar{P}^k : x = C^k y + (k-1), \quad (8.19)$$

with coefficient

$$C^k = C^k(c_1, c_2, \dots, c_{k-1}, c_{k+1}, \dots, c_N) = \sum_{j=1}^{k-1} \underbrace{(j-k)}_{<0} c_j + \sum_{j=k+1}^N \underbrace{(j-k)}_{>0} c_j, \quad (8.20)$$

and

$$\begin{cases} C^1 = C^1(c_2, c_3, \dots, c_N) = \sum_{j=2}^N (j-1)c_j, \\ C^N = C^N(c_1, c_2, \dots, c_{N-1}) = \sum_{j=1}^{N-1} (j-N)c_j. \end{cases} \quad (8.21)$$

In the above decomposition, the first sum has negative terms and the second positive terms. Further, it is clear that C^k is *independent of* c_k . The line \bar{P}^k having slope $1/C^k$ is on the point $(k-1, 0)$. Each point on \bar{P}^k is the $\bar{\pi}_0$ for a hyperplane π .

The properties and crucial role of the P^k in the development are better understood in terms of the underlying geometry. For a plane $\pi \subset \mathbb{R}^3$ as in Fig. 8.6, note that $P^1, P^2, P^3 \in \ell_\pi = \pi \cap \pi_1^s$ with $x_1(P^1) = 0, x_2(P^2) = 0, x_3(P^3) = 0$, and $\bar{\ell}_\pi = \bar{\pi}_{123}$. Recall that the coefficient vector $\mathbf{c} = (c_1, c_2, c_3)$ is normal to the plane π . Translation of $\bar{\ell}_\pi$ along \bar{P}^1 corresponds to a rotation that leaves c_2 and c_3 invariant, since C^1 is a constant, while c^1 varies. The situation is illustrated in Fig. 8.7. On P_1 , the normal vector ranges from $\mathbf{c} = (c_1^-, c_2, c_3)$ to $\mathbf{c} = (c_1^+, c_2, c_3)$, tracing a plane π_{c_1} orthogonal to all the planes with such \mathbf{c} . These planes π rotate about the axis n_1 through P_1 and normal to π_{c_1} with π intersecting the superplane

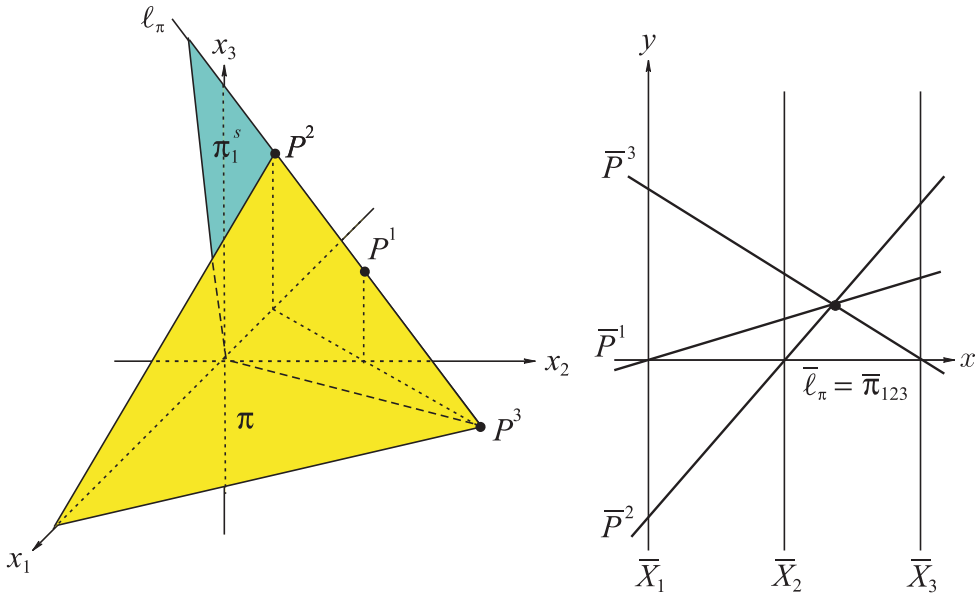


Figure 8.6. A plane π and its intersection $\ell_\pi = \bar{\pi}_{123}$ with the first superplane π_1^s . The points $P^k = \ell_\pi \cap x_i x_j$ plane for $k \neq i, j$.

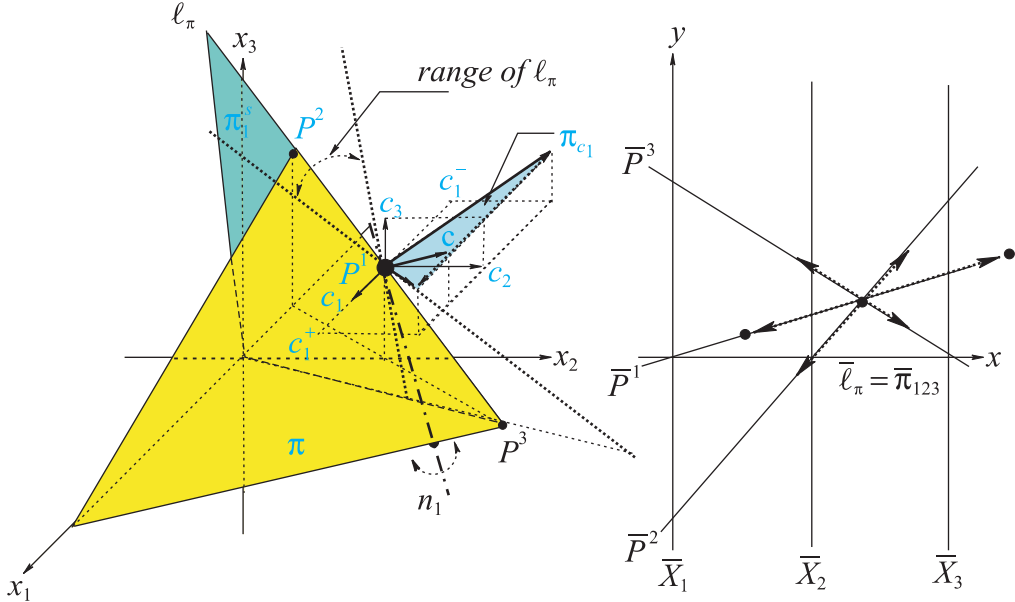


Figure 8.7. The vector $\mathbf{c} = (c_1, c_2, c_3)$ is normal to π .

On P_1 , varying just $c_1 \in [c_1^-, c_1^+]$ generates the plane π_{c_1} where \mathbf{c} ranges. The translation of $\bar{\ell}_\pi = \bar{\pi}_{123} \in \bar{P}^1$ (right) corresponds to the rotation of π about the axis $n_1 \perp \pi_{c_1}$ through P_1 . The translations along \bar{P}^2, \bar{P}^3 leaving c_2, c_3 invariant respectively correspond to similar rotations about lines on P^2 and P^3 .

π_1^s at different lines ℓ_π corresponding to different points $\bar{\ell}_\pi = \bar{\pi}_{123}$ translated along \bar{P}_1 ; clearly $n_1 \not\subset \pi_1^s$. The same considerations apply for P_2 and P_3 . The generalization to \mathbb{R}^N is direct, with the translation of $\bar{\ell}_\pi$ along \bar{P}^k corresponding to an equivalent rotation with C^k , which determines \bar{P}^k (and hence P^k), being independent of c_k .

Let

$$\left\{ \begin{array}{ll} C_m^k = C^k(c_1^+, \dots, c_{k-1}^+, c_{k+1}^-, \dots, c_N^-), \\ C_M^k = C^k(c_1^-, \dots, c_{k-1}^-, c_{k+1}^+, \dots, c_N^+), \\ C_m^1 = C^1(c_2^-, \dots, c_N^-), & C_m^N = C^N(c_1^+, \dots, c_{N-1}^+), \\ C_M^1 = C^1(c_2^+, \dots, c_N^+), & C_M^N = C^N(c_1^-, \dots, c_{N-1}^-). \end{array} \right. \quad (8.22)$$

Since $c_j^- \leq c_j \leq c_j^+$,

$$\left. \begin{aligned} \sum_{j=1}^{k-1} (j-k)c_j^- &\geq \sum_{j=1}^{k-1} (j-k)c_j \geq \sum_{j=1}^{k-1} (j-k)Wic_j^+ \\ \sum_{j=k+1}^N (j-k)c_j^+ &\geq \sum_{j=k+1}^N (j-k)c_j \geq \sum_{j=k+1}^N (j-k)c_j^- \end{aligned} \right\} \Rightarrow C_m^k \leq C^k \leq C_M^k. \quad (8.23)$$

That is, C_m^k and C_M^k are the minimum and maximum values of C^k .

The multivariate function f_N can be analyzed one variable at a time. Specifically, for c_k , the region of interest is swept by the *rotation* of the line \bar{P}^k about the point $(k-1, 0)$ between the lines

$$\begin{cases} \bar{P}_m^k : x = C_m^k y + (k-1), \\ \bar{P}_M^k : x = C_M^k y + (k-1), \end{cases} \quad (8.24)$$

from the minimum $1/C_M^k$ to maximum $1/C_m^k$ slopes. Caution: the slopes are the *reciprocals* of the C^k . The variation of c_k is due to this rotation, whose dual is a translation of P^k but *within the superplane* π_1^s , so that its representation \bar{P}^k remains a line (rather than a polygonal line).

For \mathbb{R}^3 , translations of the three P^k within the segments $[P_\ell^k, P_h^k] \subset \pi_1^s$ are indicated on the left part of Fig. 8.8. The subscripts ℓ, h are used for the endpoints P_m^k or P_M^k instead of m, M , since their relative positions depend on the signs of C_m^k, C_M^k , and this is irrelevant to the discussion now. In \parallel -coords (see the right part of Fig. 8.8) the intersection of the three sectorial regions swept by the dual rotations about the points $(k-1, 0)$ contains the points $\bar{\pi}_{O'}$ corresponding to the planes generated by the simultaneous translations of the P^k , giving us a first glimpse at a region Ω and explaining the hexagonal patterns seen in Fig. 10.43. All this begs the question: are there equivalences between translations – rotations? This is answered with Fig. 8.9 showing translation – rotation duals. Starting then from a plane π , a family of “close” planes in a neighborhood $\mathcal{N}_{O'}$ can be generated from π by the aforementioned rotations and/or the translations of the P^k shown in the picture. In due course, vertical translations for an arbitrary c_0 in π ’s equation will be added. The technical details that follow do not obscure our understanding when the underlying geometry is kept clearly mind.

8.3.4 The Region Ω

Components and Structure of Ω

After the short and incomplete preview in \mathbb{R}^3 , we proceed with the generalization to \mathbb{R}^N . Let us denote the “fan” regions swept between \bar{P}_M^k and \bar{P}_m^k by Ω_k . Now

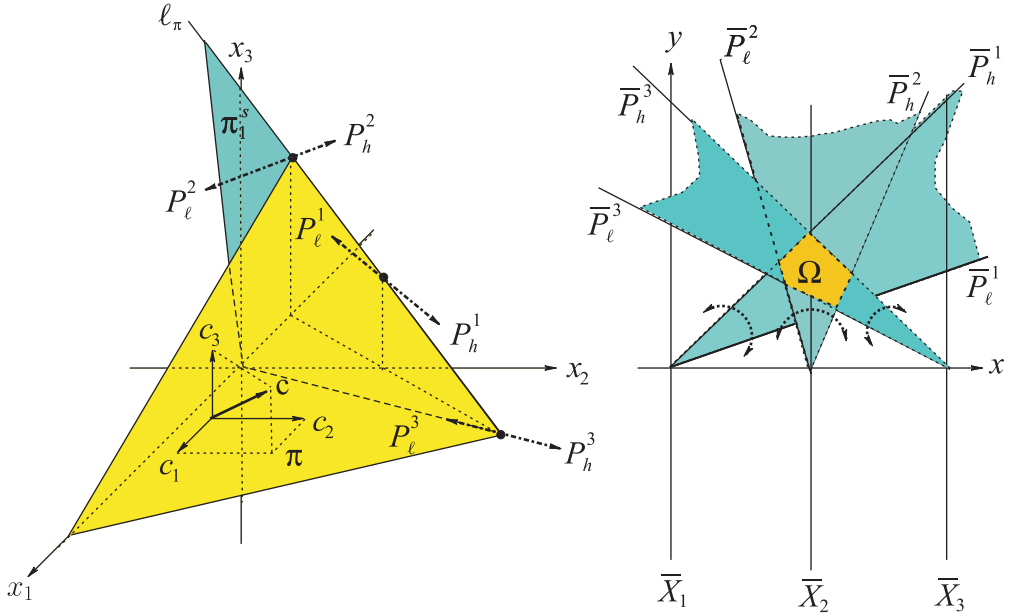


Figure 8.8. Rotation of \bar{P}^k about $(k-1, 0)$.

It is dual to the translation of P^k within a segment $[P_\ell^k, P_h^k] \subset \pi_1^s$, so that \bar{P}^k remains a straight (rather than polygonal) line. The intersection of the three sectorial regions on the right is the region Ω .

it is imperative, in order to cover all possibilities, to examine the positions of P_m^k relative to P_M^k .

Case 1. $0 < C_m^k$; then $1/C_m^k \geq 1/C^k \geq 1/C_M^k$, and \bar{P}^k ranges between the lines \bar{P}_M^k with minimum and \bar{P}_m^k maximum slopes respectively; first part of Fig. 8.10.

Case 2. $C_m^k < 0 < C_M^k$; then $1/C_m^k < 0 < 1/C_M^k$, and \bar{P}^k ranges between the lines \bar{P}_m^k with minimum and \bar{P}_M^k maximum slopes respectively. Due to the sign change, C^k goes through zero when \bar{P} is vertical at $(k-1, 0)$, corresponding to P being an ideal point and therefore π being parallel to the x_k axis; second part of Fig. 8.10.

Case 3. Finally, when $C_M^k < 0$, $0 > 1/C_m^k > 1/C^k > 1/C_M^k$, and \bar{P}^k ranges between \bar{P}_M^k with the minimum and \bar{P}_m^k with maximum slopes respectively; third part of Fig. 8.10.

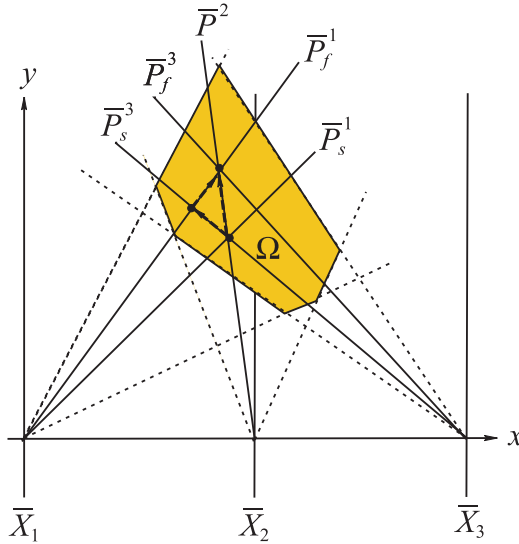


Figure 8.9. Within Ω a translation along \bar{P}^2 of $\bar{\pi}_0$ to a new position. It can also be achieved by two translations, first along a \bar{P}_s^3 then along a \bar{P}_f^1 , or equivalently by two rotations $\bar{P}_s^1 \mapsto \bar{P}_f^1$ followed by $\bar{P}_s^3 \mapsto \bar{P}_f^3$.

For the case $S(\mathbf{c}) = 0$, this analysis breaks down, since $y = 1/S$ is no longer defined, necessitating the consideration of f_N (8.49) in homogeneous coordinates.

Lemma 8.3.9. ♣ FT-3 $\Omega = \cap_{k=1}^N \Omega_k$. Either

- (a) Ω is a convex set strictly above or below the x axis, or
- (b) Ω consists of a gh region, the upper branch above and lower branch below the x axis.

**** Proof.** Each Ω_k consists of two sectorial (and of course convex) regions symmetric about the point $(k - 1, 0)$. When the intersection of the portions below the x axis is empty (as in Fig. 8.8, for example), then Ω is above the x axis and is convex, since it is the intersection of convex regions. Similarly, it may happen that all the Ω_k intersect below the x axis. Otherwise, some of the Ω_k have a nonempty intersection that is convex above, and others may have a convex intersection below the x axis. ■

In the second case, it will be seen that there exists a combination of the c_j for which $S(\mathbf{c}) = 0$, resulting in Ω being a gh . For the construction of Ω we take our

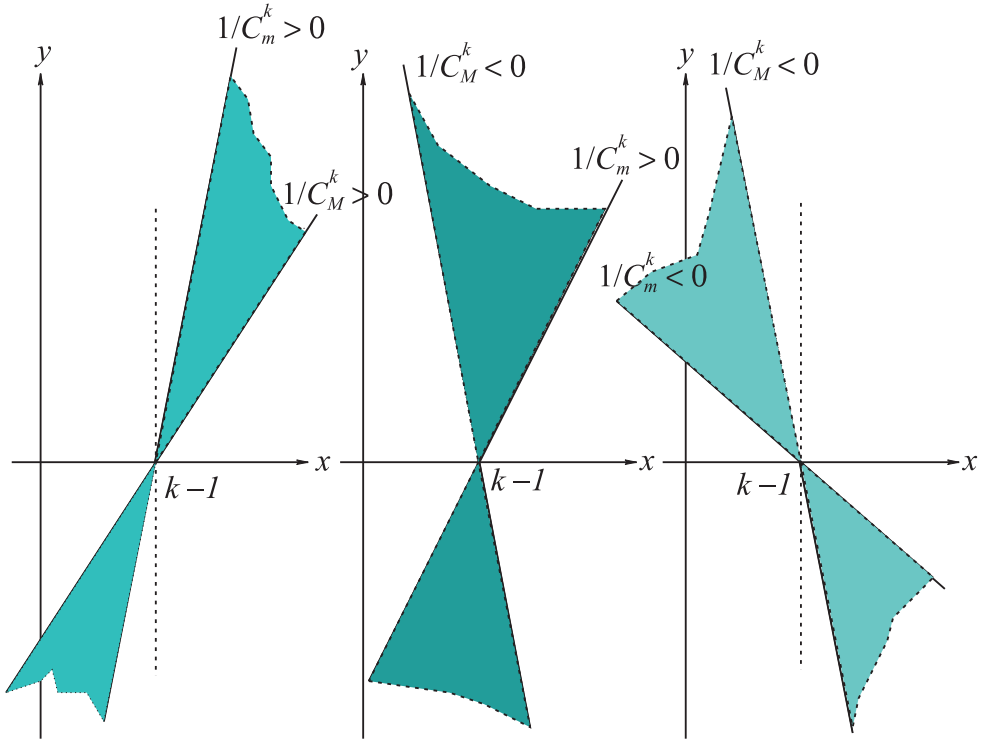


Figure 8.10. “Fan” regions Ω_k swept by the rotations of \bar{P}^k about the point $(k-1, 0)$.

cue from Fig. 8.8 and investigate the intersections $\bar{P}_m^k \cap \bar{P}_m^{k+1}$ and $\bar{P}_M^k \cap \bar{P}_M^{k+1}$ starting with the relations between the successive C_m^k and C_M^k :

$$\begin{aligned}
 C_m^{(k-1)} &= \sum_{j=1}^{k-2} (j-k+1)c_j^+ + \sum_{j=k}^N (j-k+1)c_j^- \\
 &= \sum_{j=1}^{k-2} (j-k)c_j^+ + \sum_{j=k}^N (j-k)c_j^- + \left(\sum_{j=1}^{k-2} c_j^+ + \sum_{j=k}^N c_j^- \right) \\
 &= \sum_{j=1}^{k-1} (j-k)c_j^+ + \sum_{j=k+1}^N (j-k)c_j^- + \left(\sum_{j=1}^{k-1} c_j^+ + \sum_{j=k}^N c_j^- \right).
 \end{aligned}$$

Carrying out similarly for C_M^k and rewriting,

$$\begin{cases} C_m^{(k-1)} = C_m^k + \left(\sum_{j=1}^{k-1} c_j^+ + \sum_{j=k}^N c_j^- \right), \\ C_M^{(k-1)} = C_M^k + \left(\sum_{j=1}^{k-1} c_j^- + \sum_{j=k}^N c_j^+ \right), \end{cases} \quad (8.25)$$

which points out the key vertices on the boundary ∂B :

$$\begin{cases} A_k = (c_1^+, c_2^+, \dots, c_{k-1}^+, c_k^-, c_{k+1}^-, \dots, c_N^-), \\ B_k = (c_1^-, c_2^-, \dots, c_{k-1}^-, c_k^+, c_{k+1}^+, \dots, c_N^+), \end{cases} \quad \text{with} \quad (8.26)$$

$$\begin{cases} A_1 = (c_1^-, \dots, c_N^-) = B_N, \\ B_1 = (c_1^+, \dots, c_N^+) = A_N, \end{cases}$$

so that (8.25) becomes

$$\begin{cases} C_m^{(k-1)} = C_m^k + S(A_k), \\ C_M^{(k-1)} = C_M^k + S(B_k). \end{cases} \quad (8.27)$$

These $2N$ vertices are characterized by a *single* change of sign in the superscripts at the k th position. Pursuing the cue in Fig. 8.8, we obtain the successive line intersections $\bar{P}_m^k \cap \bar{P}_m^{(k-1)}$ by

$$C_m^k y + (k-1) = C_m^{(k-1)} y + (k-2) \quad \Rightarrow \quad y = \frac{1}{S(A_k)}, \quad (8.28)$$

and correspondingly for \bar{P}_M^k , so that for $k = 2, \dots, N-1$,

$$\bar{a}_m^k = \bar{P}_m^k \cap \bar{P}_m^{(k-1)} = f_N(A_k) = (x(A_k), y(A_k)) = \left(\frac{C_m^k}{S(A_k)} + (k-1), \frac{1}{S(A_k)} \right), \quad (8.29)$$

$$\bar{a}_M^k = \bar{P}_M^k \cap \bar{P}_M^{(k-1)} = f_N(B_k) = (x(B_k), y(B_k)) = \left(\frac{C_M^k}{S(B_k)} + (k-1), \frac{1}{S(B_k)} \right). \quad (8.30)$$

For $k = 1$ and $k = N$,

$$\bar{P}_m^1 \cap \bar{P}_M^N = f(A_1) = f(B_N), \quad \bar{P}_m^N \cap \bar{P}_M^1 = f(B_1) = f(A_N). \quad (8.31)$$

Other nice and useful relations for $1 \leq r < k$ are

$$C_m^r - C_m^k = C_m^r - C_m^{(r+1)} + C_m^{(r+1)} - \dots + C_m^{(k-1)} - C_m^k = \sum_{j=r+1}^k S(A_j), \quad (8.32)$$

$$C_M^r - C_M^k = C_M^r - C_M^{(r+1)} + C_M^{(r+1)} - \dots + C_M^{(k-1)} - C_M^k = \sum_{j=r+1}^k S(B_j). \quad (8.33)$$

From these we can find the intersection points

$$\bar{a}_m^{rk} = \bar{P}_m^r \cap \bar{P}_m^{(k-1)}, \quad y(\bar{a}_m^{rk}) = \frac{(k-r)}{\sum_{j=r+1}^k S(A_j)}, \quad (8.34)$$

$$\bar{a}_M^{rk} = \bar{P}_M^r \cap \bar{P}_M^{(k-1)}, \quad y(\bar{a}_M^{rk}) = \frac{(k-r)}{\sum_{j=r+1}^k S(B_j)}, \quad (8.35)$$

and specifically for $r = 1$,

$$\bar{a}_m^{1k} = \bar{P}_m^1 \cap \bar{P}_m^{(k-1)}, \quad y(\bar{a}_m^{1k}) = \frac{(k-1)}{\sum_{j=2}^k S(A_j)}, \quad (8.36)$$

$$\bar{b}_M^{rk} = \bar{P}_M^r \cap \bar{P}_M^{(k-1)}, \quad y(\bar{b}_M^{rk}) = \frac{(k-1)}{\sum_{j=2}^k S(B_j)}. \quad (8.37)$$

There are also some important monotonicities with $S(B_1) \geq S(A_1)$ and

$$\left. \begin{aligned} S(A_k) &= \sum_{j=1}^{k-1} c_j^+ + \sum_{j=k}^N c_j^- \\ S(A_{k-1}) &= \sum_{j=1}^{k-2} c_j^+ + \sum_{j=k-1}^N c_j^- \end{aligned} \right\} \Rightarrow S(A_{k-1}) \leq S(A_k), \quad (8.38)$$

showing that

$$\frac{1}{S(A_N)} = y(A_N) \leq y(A_{N-1}) \leq \dots \leq y(A_1) = \frac{1}{S(A_1)}.$$

Similarly, $S(B_{k-1}) \geq S(B_k) \Rightarrow$

$$\frac{1}{S(B_N)} = y(B_N) \geq y(B_{N-1}) \geq \dots \geq y(B_1) = \frac{1}{S(B_1)}.$$

If there is a combination of the coefficients \mathbf{c} such that $S(\mathbf{c}) = c_1 + c_2 + \cdots + c_N = 0$, while these c_k are in $[c_k^-, c_k^+]$, then $y(A_k)$ and $y(A_{k+1})$ have different signs, so that the above monotonicities become

$$\begin{aligned} y(A_k) &\leq y(A_{k-1}) \leq \cdots \leq y(A_2) \leq y(A_1) < 0 < y(A_N) \\ &\leq y(A_{N-1}) \leq \cdots \leq y(A_{k+1}); \end{aligned} \quad (8.39)$$

showing that at most only one change of sign in the sequence of the $y(A_k)$ can occur. Similarly, at most one change of sign can occur in the sequence of the $y(B_k)$,

$$\begin{aligned} y(B_r) &\leq y(B_{r+1}) \leq \cdots \leq y(B_N) = y(A_1) < 0 < y(B_1) \\ &= y(A_N) \leq \cdots \leq y(B_{r-1}). \end{aligned} \quad (8.40)$$

8.3.5 ** Construction of Ω

Consider the first part of the lemma and specifically when Ω is above the x -axis. All values of $y(B)$ are greater than 0 for any point $B \in \mathbf{B}$, and hence $y(A_k) > 0$, $y(B_K) > 0$, so that $C_m^{(k-1)} > C_m^k$ and $C_M^{(k-1)} > C_M^k$ from (8.25). There are three possible variations for the slopes of \bar{P}_m^k, \bar{P}_M^k :

1. all slopes positive,

$$\frac{1}{C_m^N} \geq \frac{1}{C_m^{(N-1)}} \geq \cdots \geq \frac{1}{C_m^1} > 0, \quad \frac{1}{C_M^N} \geq \frac{1}{C_M^{(N-1)}} \geq \cdots \geq \frac{1}{C_M^1} > 0; \quad (8.41)$$

2. slopes change sign i.e. $C_m^1 \geq \cdots \geq C_m^k > 0 > C_m^{(k+1)} \geq \cdots \geq C_m^N$; hence

$$\begin{aligned} \frac{1}{C_m^k} \geq \cdots \geq \frac{1}{C_m^1} > 0 > \frac{1}{C_m^N} \geq \cdots \geq \frac{1}{C_m^{(k+1)}}, \quad \frac{1}{C_M^k} \\ \geq \cdots \geq \frac{1}{C_M^1} > 0 > \frac{1}{C_M^N} \geq \cdots \geq \frac{1}{C_M^{(k+1)}}; \end{aligned} \quad (8.42)$$

3. all slopes are negative,

$$0 > \frac{1}{C_m^N} \geq \frac{1}{C_m^{(N-1)}} \geq \cdots \geq \frac{1}{C_m^1}, \quad 0 > \frac{1}{C_M^N} \geq \frac{1}{C_M^{(N-1)}} \geq \cdots \geq \frac{1}{C_M^1}. \quad (8.43)$$

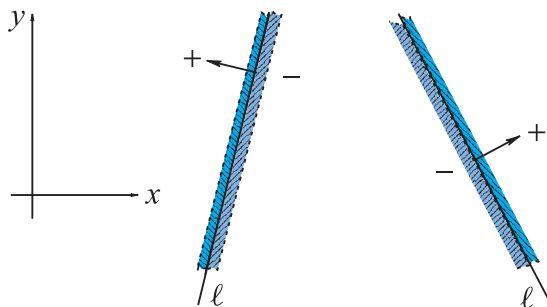


Figure 8.11. Orienting the half-spaces.

For each side of a line $\ell : ax + by = c$ for $c \geq 0$. Points in the portion “-”: $ax + by < c$ are “below” ℓ and “above” with the reverse inequality. For horizontal lines, “+” is the positive y direction, and for vertical lines, the positive x direction.

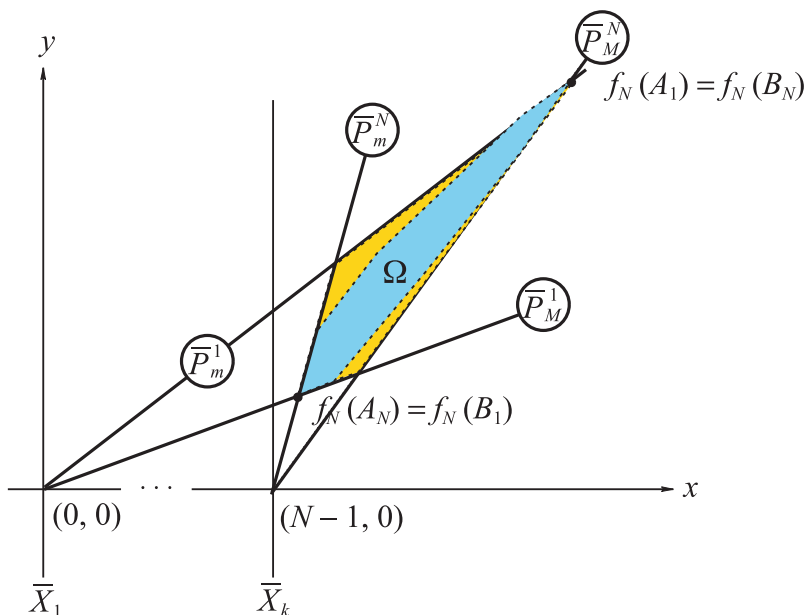


Figure 8.12. Starting the construction of Ω .

Here the slopes of the lines \bar{P}_m^k, \bar{P}_M^k are all positive; Ω is enclosed within the quadrilateral $\Omega_1 \cap \Omega_N$.

Figure 8.11 recalls the “above”/“below” relation for a point P and a line ℓ , which is helpful in what follows. Recall from Lemma 8.3.9 that $\Omega \subset \Omega_k$. The intersection $\Omega = \Omega_1 \cap \Omega_N = \Omega_{1N}$, where Ω_1 is the sectorial region between

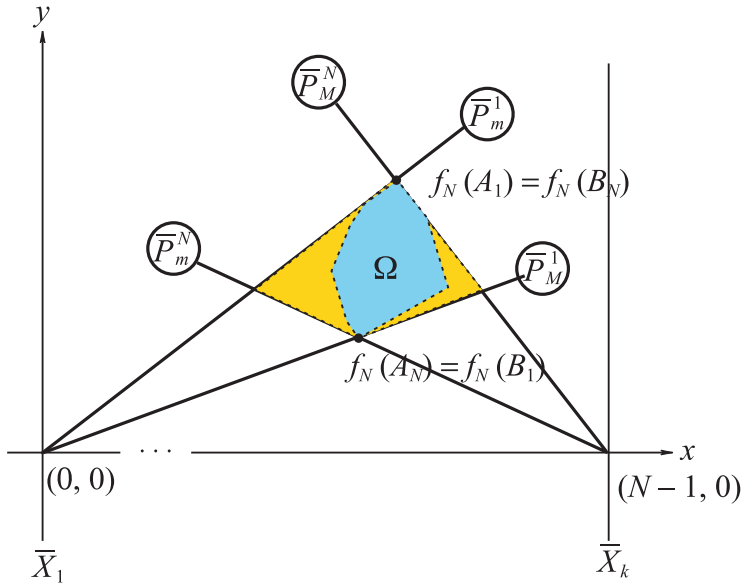


Figure 8.13. The slopes of the lines \bar{P}_m^k, \bar{P}_M^k change sign.
 Ω is enclosed within the quadrilateral $\Omega_1 \cap \Omega_N$.

\bar{P}_M^1, \bar{P}_m^1 , and Ω_N between \bar{P}_M^N, \bar{P}_m^N , is a quadrilateral with $f_N(A_1) = f_N(B_N)$, $f_N(A_N) = f_N(B_1)$ its highest and lowest vertices respectively. For case 1, shown in Fig. 8.12, all of Ω is to the right of the \bar{X}_k axis and is of course contained in the quadrilateral Ω_{1N} . For 2, shown in Fig. 8.13, Ω is between the \bar{X}_1 and \bar{X}_N axes, the slopes of the lines changing sign at some intermediate index according to (8.41). In the third case, (8.43), Ω is on the left of the \bar{X}_1 axis, as seen in Fig. 8.14. For Ω below the x axis, all the values of $y(B)$ are negative for any point $B \in \mathbf{B}$, and hence $y(A_k) < 0, y(B_K) < 0$. Then from (8.25), $C_m^{(k-1)} < C_m^k$ and $C_M^{(k-1)} < C_M^k$, and the three possibilities for the slopes of \bar{P}_m^k, \bar{P}_M^k analogous to (8.41), (8.42), (8.43) result in Ω being the mirror images, with respect to the x axis, of the previous situation, with $f_N(A_1)$ still higher than $f_N(A_N)$ as in (8.39), (8.40).

It is clear that the vertices $f_N(A_1), f_N(A_N)$ together with a segment from each of \bar{P}_m^1, \bar{P}_M^N belong to $\partial\Omega$. Therefore, $f_N(A_1), f_N(A_N) \in \Omega_k$ for any k . Hence, as illustrated in Fig. 8.15, neither of the lines \bar{P}_m^k, \bar{P}_M^k intersects the line joining $f_N(A_1), f_N(A_N)$, for if one of them did, then $f_N(A_1), f_N(A_N)$ would lie on opposite sides of that line, so that *both* of them cannot belong to $\partial\Omega$. At this stage, the order of the \bar{P}_m^k, \bar{P}_M^k is immaterial, but for the sake of the discussion we take the

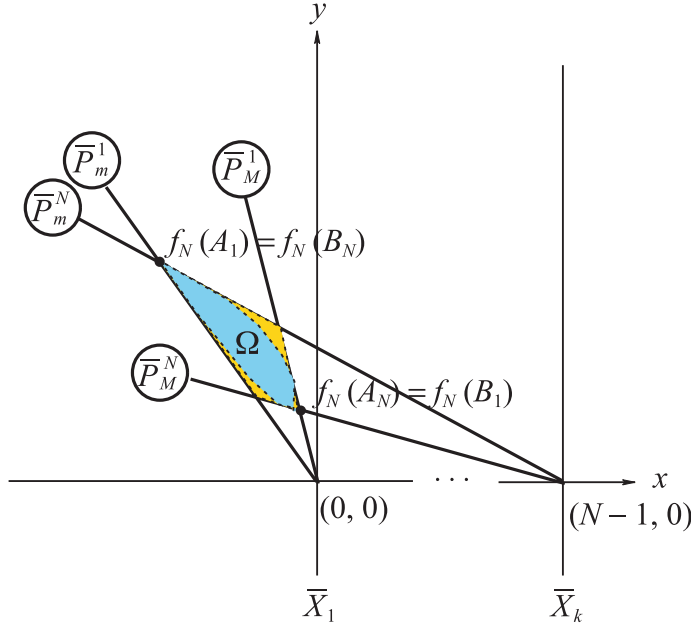


Figure 8.14. The slopes of the lines \bar{P}_m^k, \bar{P}_M^k are all negative. Ω is enclosed within the quadrilateral $\Omega_1 \cap \Omega_N$.

arrangement shown with $\bar{n}_{ik} = \bar{P}_M^k \cap \bar{P}_m^1$ higher in y than $y(A_1)$. Hence $\bar{n}_{ik} \notin \partial\Omega$, meaning that the segment of \bar{P}_m^1 belonging to $\partial\Omega$ has an endpoint $\bar{a}_{ik} = \bar{P}_m^1 \cap \bar{P}_m^k$ for some k . By the same argument, there is an endpoint $\bar{b}_{ik} = \bar{P}_M^1 \cap \bar{P}_M^k \in \partial\Omega$. From (8.36) and (8.39), a comparison of the y coordinates of the points $\bar{a}_{12} = f_N(A_2)$ and \bar{a}_{1k} yields

$$y(\bar{a}_{1k}) = \frac{(k-1)}{\sum_{j=2}^k S(A_j)} \leq \frac{(k-1)}{\sum_{j=2}^k S(A_2)} = \frac{1}{S(A_2)} = y(A_2), \quad (8.44)$$

illustrated in Fig. 8.16, showing that the point \bar{a}_{12} is $f_N(A_2) \in \partial\Omega$ with $\bar{a}_{1k} \notin \Omega$ for $k \geq 3$. So it is the segment with endpoints $f_N(A_1), f_N(A_2)$ that belongs to $\partial\Omega$, and similarly for the vertices $f_N(B_N), f_N(B_{N-1}) \in \partial\Omega$ together with the segment joining them.

Let A_k, A_{k+1} be the endpoints of the edge a_k , and B_k, B_{k+1} the endpoints of the edge b_k , respectively, on ∂B . Then

$$\begin{cases} a_k = (c_1^+, c_2^+, \dots, c_{k-1}^+, c_k, c_{k+1}^-, \dots, c_N^-), \\ b_k = (c_1^-, c_2^-, \dots, c_{k-1}^-, c_k, c_{k+1}^+, \dots, c_N^+), \end{cases} \quad c_k \in [c_k^-, c_k^+]. \quad (8.45)$$

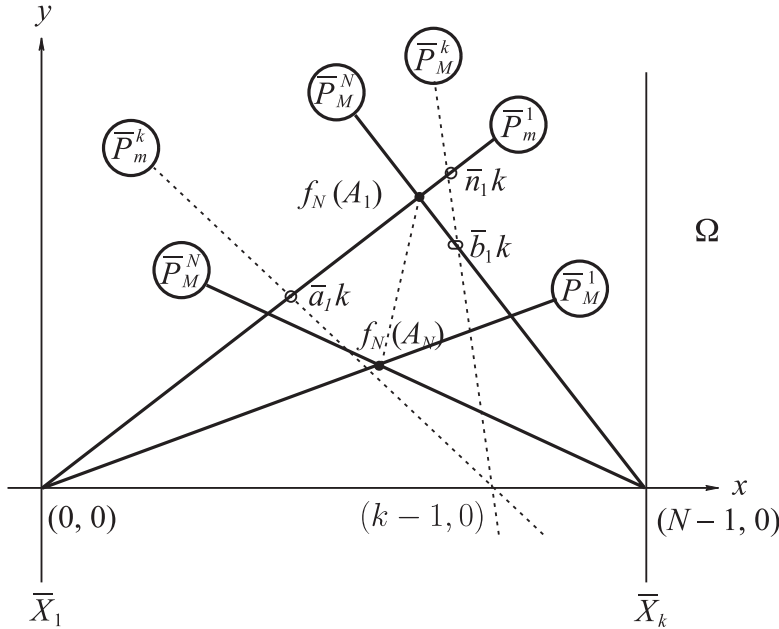


Figure 8.15. The segment connecting $f_N(A_1)$ and $f_N(A_N)$ is contained in Ω_k for all Ω_k .

The point \bar{n}_{ik} is equal to $\bar{P}_m^1 \cap \bar{P}_M^k \notin \partial\Omega$ for $k \geq 2$, but the point \bar{a}_{1k} equals $\bar{P}_m^1 \cap \bar{P}_m^k \in \partial\Omega$ for some k .

Restating, we have established that the image $f_N(a_1)$ of the edge a_1 joining A_1, A_2 , and similarly the edge $f_N(b_1)$ with endpoints $f_N(B_N), f_N(B_{N-1})$, belongs to Ω (recall that $A_1 = B_N$). Inductively, the same argument shows that $f_N(a_k), f_N(b_k) \subset \partial\Omega \forall k = 1, \dots, N$. We have shown that the construction of Ω is accomplished by sweeping along the x axis to the points $(0, 0), \dots, (k-1, 0), \dots, (N-1, 0)$ where the sectorial regions Ω_k are anchored, intersecting them sequentially, $(\Omega_1 \cap \Omega_2) \cap \dots \cap \Omega_k$, terminating with Ω_N to generate Ω , as illustrated in Fig. 8.18.

The three possibilities for the segments forming $\partial\Omega$ are shown in Fig. 8.17, the third case (on the right) still to be discussed. In due course it will be seen that the segments can be either semi-infinite lines or projective complements of finite segments. Equivalently, the construction of Ω corresponds to the judicious traversal of the portion of ∂B consisting of the $2N$ vertices A_k, B_k , specified in (8.45), on the path \mathcal{P} along the $2N$ edges $a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N$. At each vertex, say A_k , the point $f_N(A_k)$ in the xy plane is computed and joined by a straight line segment to the previous one $f_N(A_{k-1})$. To emphasize, out of a total of 2^N vertices (and edges) on ∂B , only $2N$ are needed for the construction of Ω . The traversal

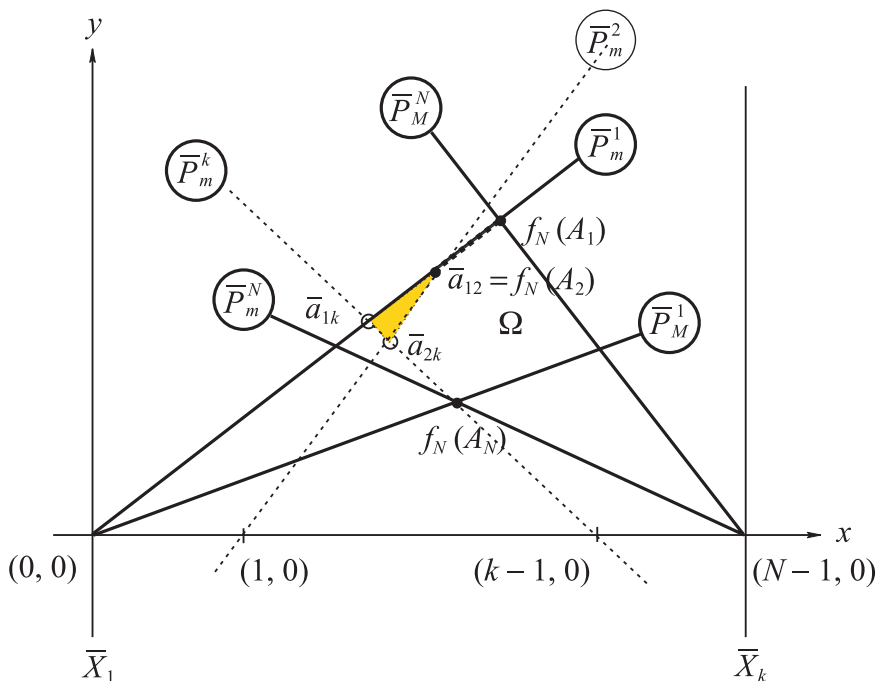


Figure 8.16. The segment of \bar{P}_m^1 with endpoints $f_N(A_1)$, $f_N(A_2)$. It is included in $\partial\Omega$, with the triangular region shown excluded from Ω .

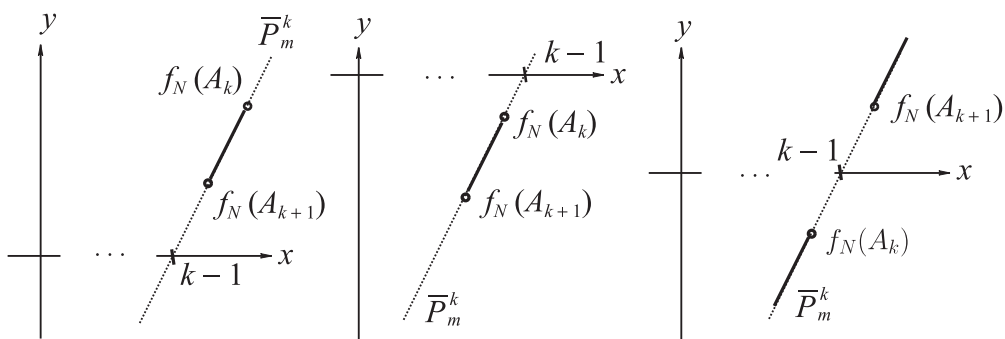


Figure 8.17. The segments $f_N(a_k) \subset \bar{P}_m^k$.

This is for the case where Ω is completely above or below the x axis (left and center) and the split on the right, which can occur at most for one k . The situation for the edges $f_N(b_k) \subset \partial\Omega$ is the same. The segments in the first two cases can also be semi-infinite lines, as in Fig. 8.22.

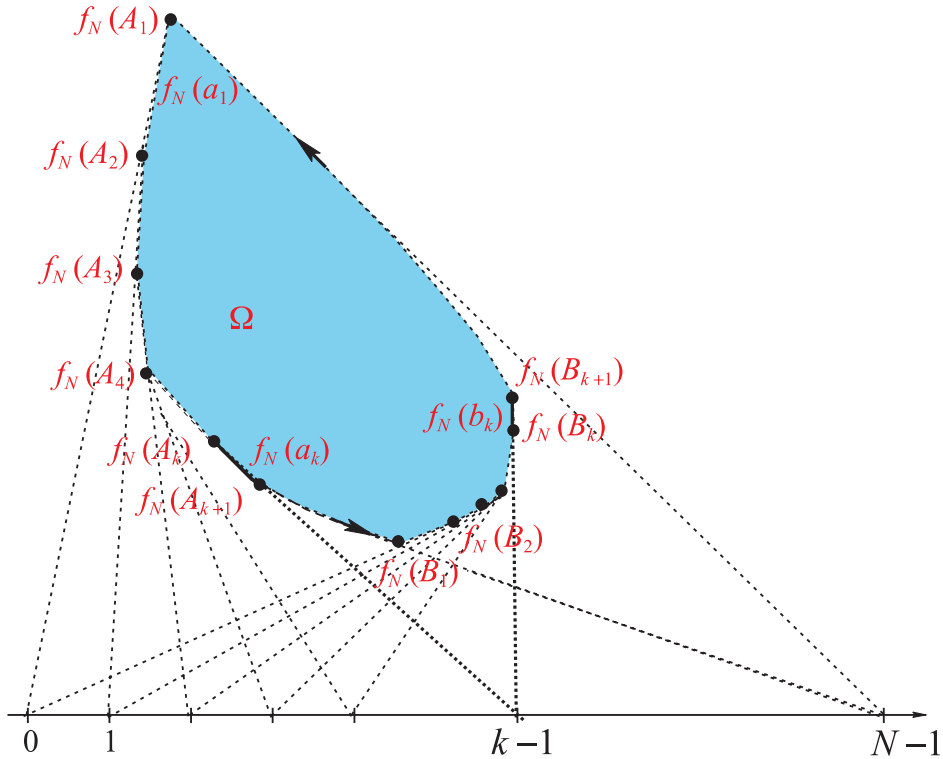


Figure 8.18. ♣ FT-4 Construction of Ω .

The k th region Ω_k contributes the vertices $f_N(A_k)$, $f_N(B_k)$ to $\partial\Omega$, and together with the region Ω_{k+1} , the vertices $f_N(A_{k+1})$, $f_N(B_{k+1})$ and edges $f_N(a_k)$, $f_N(b_k)$, as highlighted.

direction for a_k is $c_k^+ \rightarrow c_k^-$ and $c_k^- \rightarrow c_k^+$, for b_k , shown in Fig. 8.19 for 3-D and Fig. 8.20 for N -D:

$$\begin{array}{ccc}
 \nearrow & A_2 \longrightarrow A_3 \longrightarrow \cdots \longrightarrow A_k \longrightarrow \cdots \longrightarrow A_{N-1} & \searrow \\
 \mathcal{P} : A_1 = B_{N+1} & & B_1 = A_{N+1} \\
 \nwarrow & B_{N-1} \longleftarrow B_{N-2} \longleftarrow \cdots \longleftarrow B_r \longleftarrow \cdots \longleftarrow B_2 & \swarrow
 \end{array} \tag{8.46}$$

The construction of Ω is akin to a central problem in computational geometry: the construction of the *convex hull* of a planar set of N points. This problem is equivalent to sorting and can be accomplished with $O(N \log N)$ worst-case computational complexity [148]. In our case, the sweep shown in Fig. 8.18 provides the $2N$ vertices on $\partial\Omega$ already *sorted* according to their connectivity (order in which

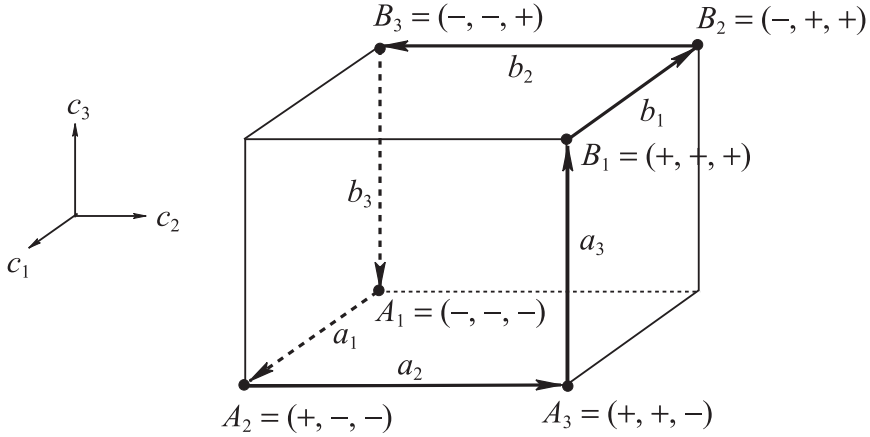


Figure 8.19. 3D-box B in the space of coefficients.

The vertices and edges, three each A_k, B_k, a_k, b_k along the path \mathcal{P} are shown. The notation \pm indicates the sign of c_k^\pm at the k th component.

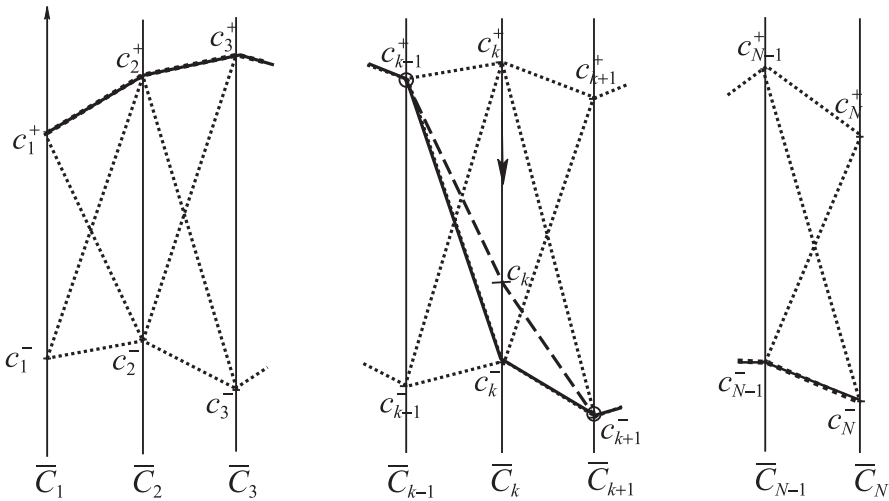


Figure 8.20. Image \bar{B} of domain B .

This is an N -dimensional box in the coefficient space $c_1 \times c_2 \times \cdots \times c_N$. The dotted lines are the polygonal lines representing the box's vertices. The solid line shows the vertex \bar{A}_k , and the dashed portion (together with the remaining solid line) shows one of the points on the edge \bar{a}_k the arrow on the \bar{C}_k axis is the direction of traversal from $c_k^+ \rightarrow c_k^-$. Each full traversal of $c_k \in [c_k^-, c_k^+]$ corresponds to an edge a_k , one of the N edges on the vertex A_k . The full path \mathcal{P} can be traced in this manner.

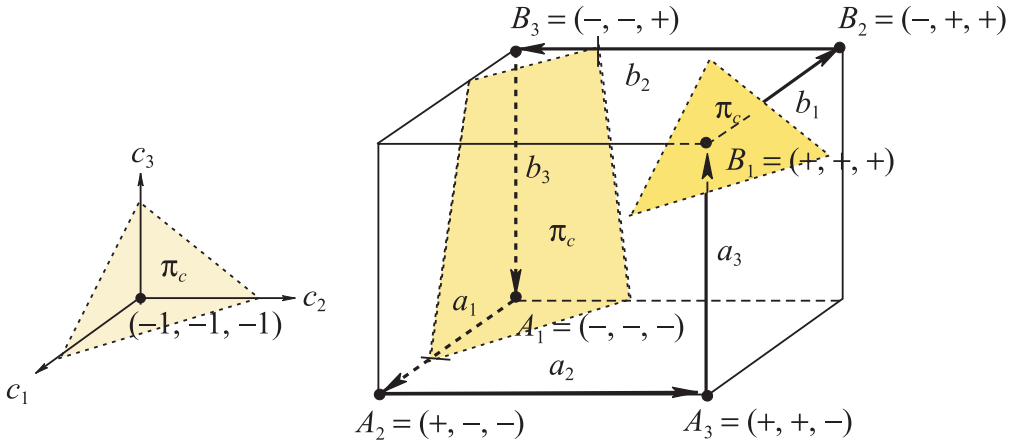


Figure 8.21. The box B in the 3-D space of coefficients c_1, c_2, c_3 , (its positions with respect to the plane $\pi_c : c_1 + c_2 + c_3 = 0$.) Namely, $\pi_c \cap B = \emptyset$, or if $\pi_c \cap B \neq \emptyset$, π_c is a supporting plane at either vertex A_1 or B_1 , or π_c intersects two edges of the path \mathcal{P} .

they are joined), that is, the path in (8.46), thus lowering the required complexity to $O(N)$.

We have been carefully skirting the prospect of the sum $S(\mathbf{c}) = \sum_{j=1}^N c_j = 0$ for some $\mathbf{c} \in B$, and this is dealt with now. In $\mathbf{c} \in \mathbb{R}^N$, the coefficient space $\pi_c : \sum_{j=1}^N c_j = 0$ is a hyperplane, and its relative position to the box B has important ramifications with respect to the corresponding Ω . So for the determination of Ω , the procedure is to first check

$$\text{sign}(S(A_1)S(B_1)) = \begin{cases} + \Rightarrow \pi_c \cap B = \emptyset & \text{or} \\ - \Rightarrow \pi_c \cap B \neq \emptyset & \text{with } S(A_1) < 0, S(B_1) > 0. \end{cases} \quad (8.47)$$

Figure 8.21 illustrates these possibilities in \mathbb{R}^3 with π_c away from B , intersecting some edges of B or containing only one of its vertices. With $\mathbf{n} = (1, 1, \dots, 1) \perp \pi_c$, clearly the only vertices for which π_c can be a supporting plane of B are A_1 or B_1 . For if π_c intersects the diagonal joining A_1 or B_1 , then it must also intersect some edges and hence cannot be a supporting plane. Drawing these three situations in \mathbb{R}^N with π_c and the N -dimensional box B in Fig. 8.20 is left as an exercise (Exercise 4). To treat the case $B \cap \pi_c \neq \emptyset$, it is necessary to consider the function f_N of (8.49) in homogeneous coordinates:

$$f_N(\mathbf{c}) = (x(\mathbf{c}), y(\mathbf{c})) = \left(\sum_{j=1}^N (j-1)c_j, 1, S(\mathbf{c}) \right). \quad (8.48)$$

Consider the case that $\pi_c \cap B$ is either A_1 or B_1 . The image of that vertex is an ideal point with slope $1/\sum_{j=1}^N (j-1)c_j$, and Ω is a *uc*, an unbounded convex set in the nomenclature of Chapter 7. The construction of $\partial\Omega$ proceeds exactly as outlined above, the exception being that Ω opens in the direction of its one ideal point, and the edges on $f_N(B_2)$ and $f_N(A_{N-1})$ are half-lines. Such an example with Ω opening downward and $f_N(B_1)$ the ideal point is shown in Fig. 8.22. Thus far, we have seen that the image of a_k is a finite segment $f_N(a_k) \subset \bar{P}_m^k$ when $\pi_c \cap B = \emptyset$ or a semi-infinite line when $\pi_c \cap B = A_1$ or B_1 , and the same holds for the images of b_k . The remaining possibility is that π_c intersects several edges of $\partial\Omega$. There are N edges at each vertex, but inspection of the path \mathcal{P} in (8.46) as well

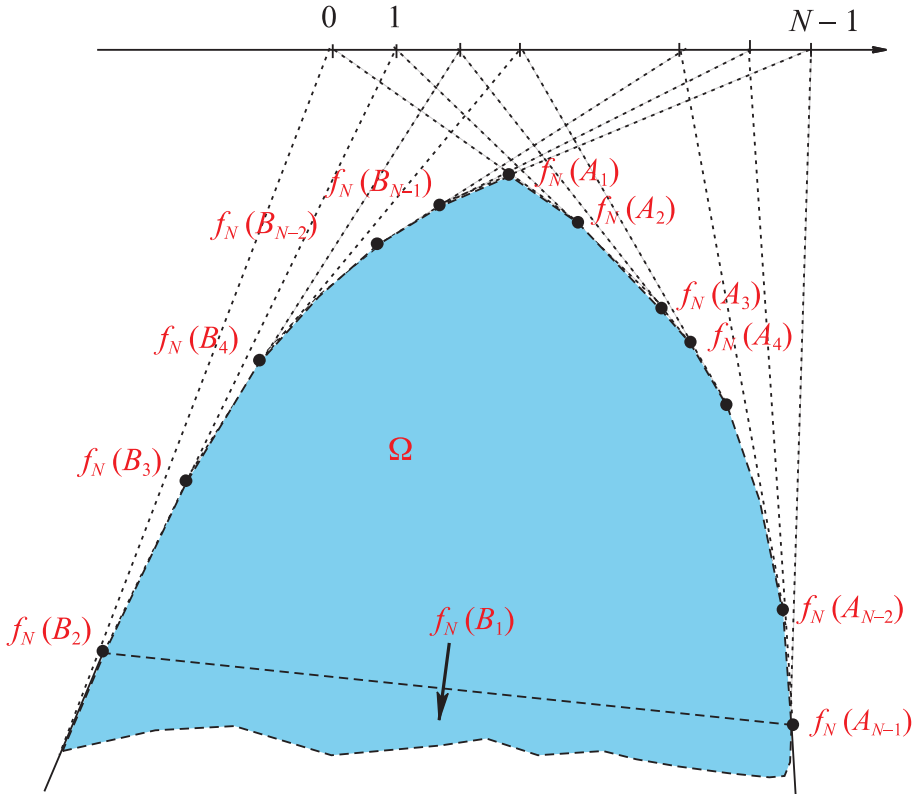


Figure 8.22. Ω can be a *uc*. Here $f_N(B_1)$ is an ideal point.

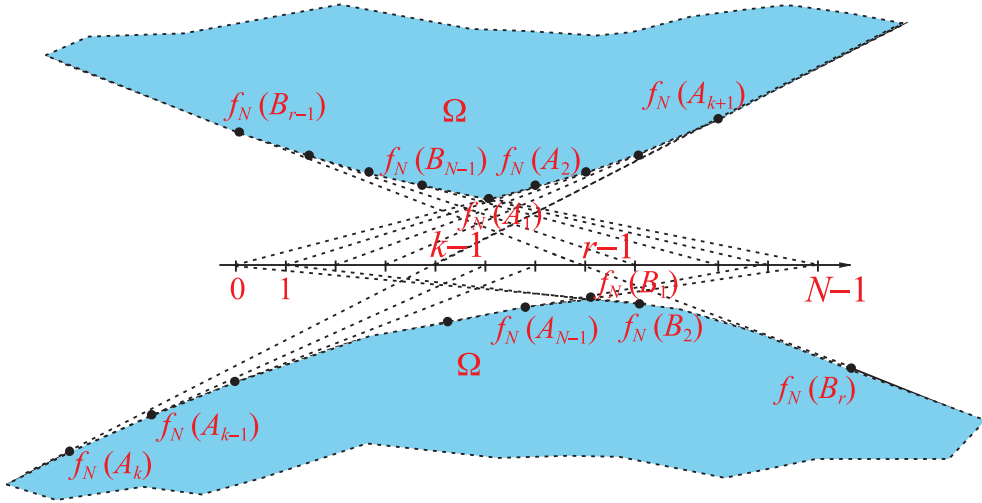


Figure 8.23. Ω can be a *gh*.

as the monotonicities in (8.39), (8.40) shows that *at most two edges of that path are intersected by π_c* ; this is also seen in Fig. 8.21. The image of such an edge, as shown on the right of Fig. 8.17, is the projective complement of the finite segment with $f_N(A_{k+1})$ now higher than $f_N(A_k)$. Then Ω is a *gh*, a *generalized hyperbola* in the nomenclature of Chapter 7 on curves as shown in Fig. 8.23, whose construction proceeds according to the order given by (8.39), (8.40). A variation occurs when π_c intersects ∂B at one of the vertices V not A_1 or B_1 . Then Ω is a *gh* with one asymptote corresponding to the ideal point \bar{V} being vertical, as shown in Fig. 7.23 and Fig. 7.30 (see Exercise 5).

For easy reference, the construction by which the region Ω is obtained is called the “*scissors algorithm*,” that is, starting with the “*fan*” regions as shown in Fig. 8.10, their intersection described in Lemma 8.3.9, and the construction as described in this section, in particular (8.41), (8.42), (8.43). Clearly this construction algorithm outputs specific types of regions, and not any $2N$ -gon region can be its output. The conditions resulting in the three types of regions are brought “under one roof” visually in Fig. 8.24 and formally below.

Theorem 8.3.10 (T. Matskewich) *The region Ω is $2N$ -gon above or below (but not intersecting) the x -axis, which is*

1. **bc** if $\pi_c \cap B = \emptyset$,
2. **uc** if $\pi_c \cap B = \{A_1\}$ or $\pi_c \cap B = \{B_1\}$,

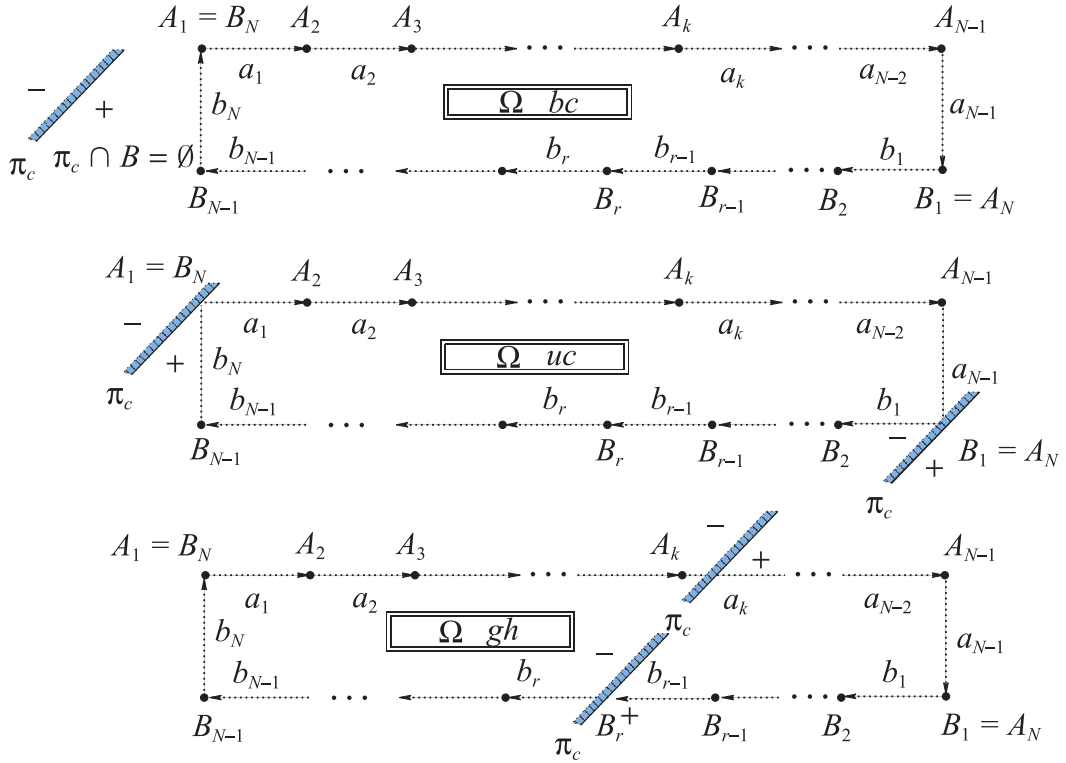


Figure 8.24. The intersection of the hyperplane π_c with the path $\mathcal{P} \subset B$ determines the type of the region Ω . When $\pi_c \cap B = \emptyset$ (top), Ω is a bc (bounded convex) $2N$ -gon as in Fig. 8.18; otherwise, a uc (unbounded convex) when π_c is a supporting hyperplane to B at either vertex A_1 or B_1 (middle) (Fig. 8.22), or as in the bottom part a gh (generalized hyperbola) (Fig. 8.23) when π_c intersects more than one segment B . In addition, if π_c cuts through one of the vertices V , then \bar{V} is an ideal point and is a vertical asymptote of Ω (Exercise 5).

3. **gh** if π_c intersects B at more than one edge, and has a vertical asymptote if in addition, π_c contains a vertex of B .

The construction of the remaining regions, Ω 's companions, forming the image of the neighborhood NH given by (8.17) is basically the same and is the next section's topic. Incidentally, NH is a complex hypersurface, the analogue of an N -Dimensional gh , and as we see, it is represented by $N - 1$ planar regions, which in turn are bc , uc , or gh . This lays the groundwork for the representation of hypersurfaces in the next chapter and shows the generality that has been achieved.

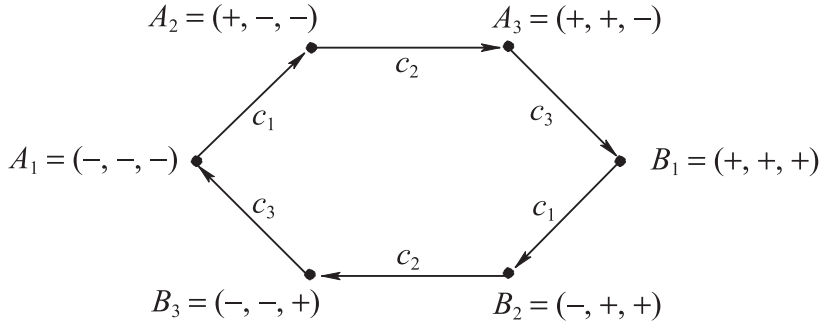


Figure 8.25. Pathway for the computation of $\partial\Omega$ when $N = 3$.

The label c_i indicates the *only* coefficient varying along the edge. Starting from $A_1 = (-, -, -)$, the next vertex $A_2 = (+, -, -)$ is found by the variation of c_1 between its extreme values. All the remaining vertices are found in the same way.

After all this preparation, let us show an example of the theorem's Case 1 for $N = 3$. What this amounts to is knowing the vertices $A_1, A_2, A_3, B_1, B_2, B_3$ for the computation of the boundary of $\Omega = \overline{NH}_O$. These are among the vertices of the 3-D box in Fig. 8.19. On the path shown in Fig. 8.25, the label c_k indicates the *only* coefficient varying along that edge. The resulting hexagon for $c_1 \in [1, 2]$, $c_2 \in [2, 3]$, $c_3 \in [3, 4]$ is drawn on the left side of Fig. 8.26 using the formula for $\bar{\pi}_O$. ♣ FT-4e

8.3.6 ** The Full Image \overline{NH}

To generate the regions $\overline{NH}_{i'}$ of \overline{NH} for $i = 0, \dots, N$ it suffices to find the vertices of $\partial\mathbf{B}$, which in turn provide the vertices of $\partial\overline{NH}_{i'}$, analogous to the traversal (8.46) for $i = 0$. The $2N$ -gon $\overline{NH}_{1'}$ contains the points $\bar{\pi}_{1'}$, which are the function values of

$$f_{N_{1'}} : \mathbb{R}^N \mapsto \mathbb{R}^2$$

in x, y coordinates, where now,

$$f_{N_{1'}}(\mathbf{c}) = (x_{1'}(\mathbf{c}), y(\mathbf{c})) = \left(\frac{Nc_1 + \sum_{j=1}^N (j-1)c_j}{S}, \frac{1}{S} \right). \quad (8.49)$$

The analysis as in Section 8.3.3 involves finding the coordinate x in terms of $y = 1/S$ via the decomposition below, for $k = 1, 2, \dots, N$:

$$x = \frac{Nc_1 + \sum_{j=1}^N (j-1)c_j}{S} = k + \frac{Nc_1 + \sum_{j=1}^N (j-k-1)c_j}{S}.$$

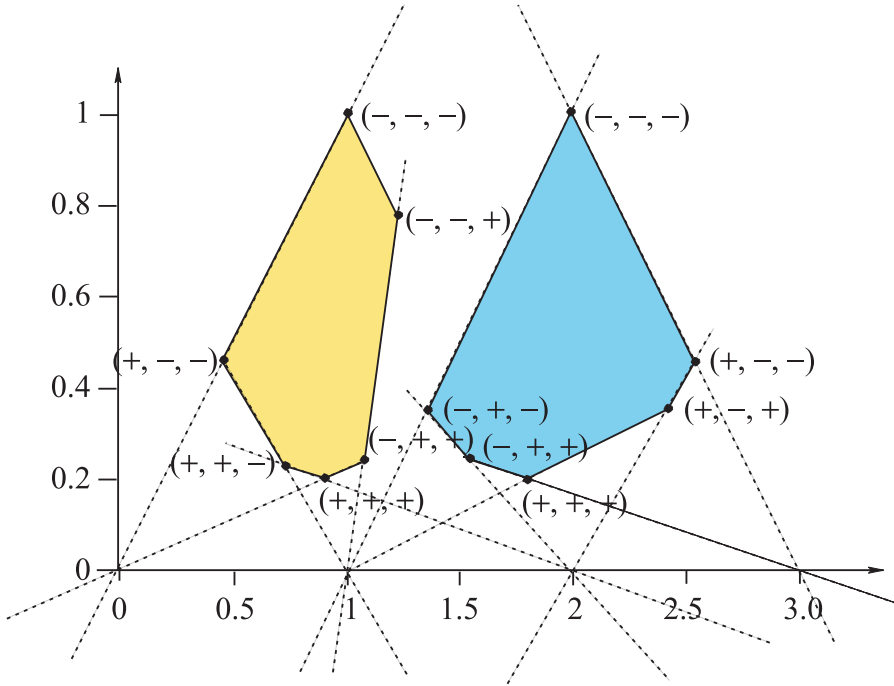


Figure 8.26. The hexagonal regions $\Omega = \overline{NH}_O'$ (left) and \overline{NH}_1' (right).

Here the family of planes is specified by $c_1 \in [1/3, 1.5]$, $c_2 \in [1/3, 2.5]$, $c_3 \in [1/3, 1]$. Compare this diagram with Fig. 10.43 at the beginning of the chapter.

And for (8.57), we obtain the linear relation

$$\bar{P}_{1'}^k : x = C_{1'}^k y + k. \quad (8.50)$$

The subscript “1'” distinguishes the above from P_0^k , which until now was denoted by P^k . The coefficient here is

$$\begin{aligned} C_{1'}^k &= C^k(c_1, c_2, \dots, c_k, c_{k+2}, \dots, c_N) \\ &= Nc_1 + \sum_{j=1}^k \underbrace{(j-k-1)}_{<0} c_j + \sum_{j=k+2}^N \underbrace{(j-k-1)}_{>0} c_j, \end{aligned} \quad (8.51)$$

with

$$\begin{cases} C_{1'}^1 = C^1(c_1, c_3, \dots, c_N) = (N-1)c_1 + \sum_{j=3}^N (j-2)c_j, \\ C_{1'}^N = C^N(c_2, \dots, c_{N-1}, c_N) = \sum_{j=2}^N (j-N-1)c_j. \end{cases} \quad (8.52)$$

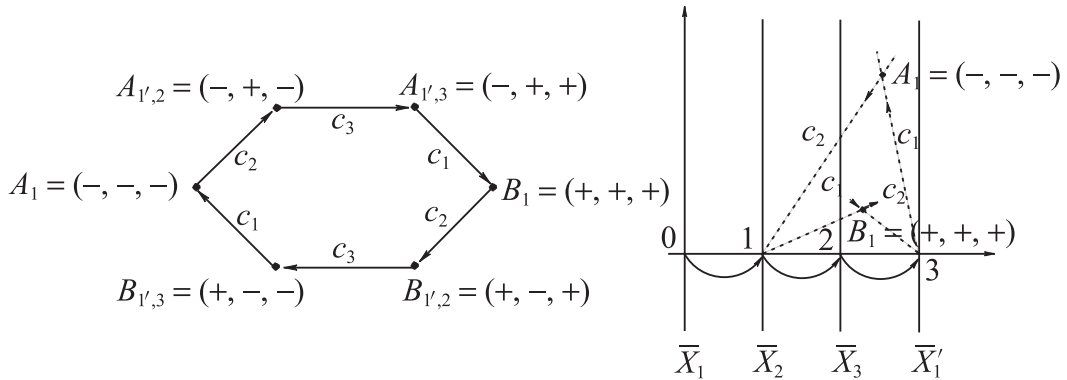


Figure 8.27. Pathway for the construction of $\partial \overline{NH}_{1'}$ for $N = 3$ (left).

These are the $\bar{\pi}_{1'}$ points based on the axis triple $\bar{X}_2, \bar{X}_3, \bar{X}_{1'}$. The labels c_i indicate the *only* coefficient varying along the edge. This traversal is found by applying the shift $c_k \rightarrow c_{k+1} \bmod 3$ to the previous diagram in Fig. 8.25 starting with $A_1 = (-, -, -)$ and c_2 varying between its extreme values along the edge connecting it to the next vertex $A_{1',2} = (-, +, -)$ (while c_1, c_3 are constant). All the remaining vertices are determined in the same way. The axis shift by 1 for the construction of $\overline{NH}_{1'}$ is shown on the right. The top and bottom vertices are indicated by the vertices A_1, B_1 of \mathbf{B} , rather than the cumbersome notation $f_{N_{1'}}(A_1), f_{N_{1'}}(B_1)$ used to compute them, while the lines are labeled by the coefficients varying along them.

The key observation is that $C_{1',m}^k$ is independent of c_{k+1} , whereas C^k is independent of c_k . So the shift $c_k \rightarrow c_{k+1} \bmod N$ applied to all the previous formulas found from C^k (henceforth denoted by $C_{0'}^k$) yields the correct relations for $C_{1'}^k$, namely,

$$\left\{ \begin{array}{l} C_{1',m}^k = C_{1'}^k(c_1^+, \dots, c_k^+, c_{k+2}^-, \dots, c_N^-), \\ C_{1',M}^k = C^k(c_1^-, \dots, c_k^-, c_{k+2}^+, \dots, c_N^+), \\ C_{1',m}^1 = C_{1'}^1(c_1^-, \dots, c_N^-), C_{1',m}^N = C_{1'}^N(c_2^+, \dots, c_N^+), \\ C_{1',M}^1 = C_{1'}^1(c_1^+, \dots, c_N^+), C_{1',M}^N = C_{1'}^N(c_2^-, \dots, c_N^-). \end{array} \right. \quad (8.53)$$

As before, the subscripts m and M indicate that $C_{1',m}^k$ and $C_{1',M}^k$ are the minimum and maximum values of $C_{1'}^k$.

By applying the shift $c_k \rightarrow c_{k+1} \bmod N$ to the previous results, we can avoid the laborious calculations to obtain the needed vertices of \mathbf{B} . The y -range of all the $\overline{NH}_{i'}$ is the same, so the vertices A_1 and B_1 , which yield top and bottom vertices, are included for all i . It is helpful to return to the 3D example and find the vertices traversed in $\partial \mathbf{B}$ for the construction of $\overline{NH}_{1'}$, as shown on the left part of Fig. 8.27. From these vertices the second hexagon shown in Fig. 8.26 is constructed using

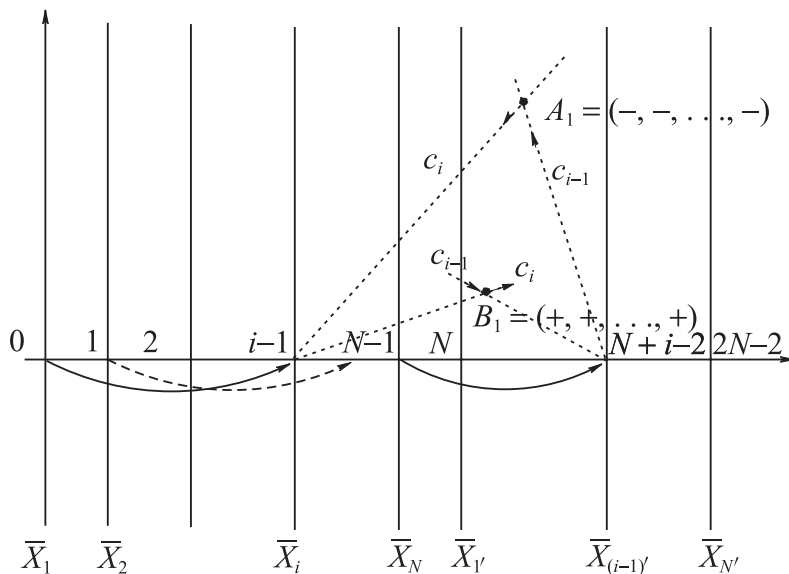


Figure 8.28. Generalizing the shift by i , $k \rightarrow k + i \bmod N$, units in \mathbb{R}^N .

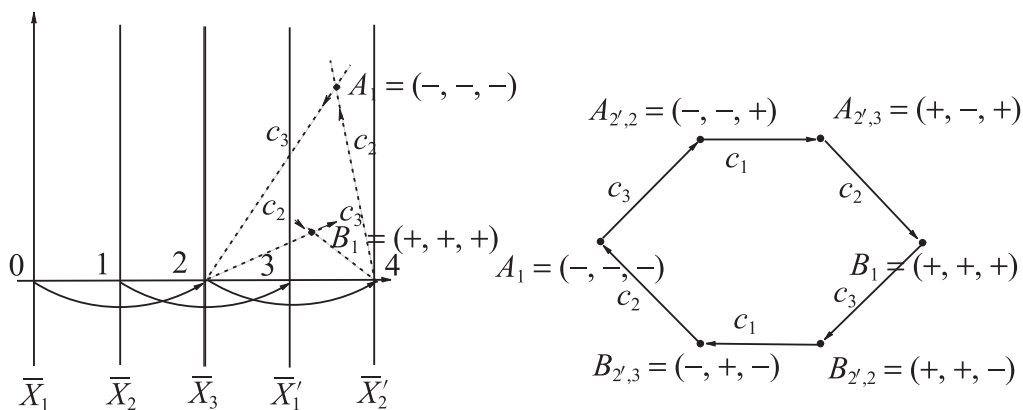


Figure 8.29. The shift by 2 units, $k \rightarrow k + 2 \bmod 3$ and the resulting path on $\partial \mathbf{B}$.

the formula for $\bar{\pi}_1$. The shift and its results are shown pictorially on the right part of Fig. 8.27. The matter easily generalizes to the shift $k \rightarrow k + i \bmod N$ units in \mathbb{R}^N to successively construct the system of N axes $\bar{X}_{i'}, \dots, \bar{X}_{(N+i-1)'}$ for the construction of the $\overline{NH}_{i'}$ as shown in Fig. 8.28. Note that the distance between the \bar{X}_i and \bar{X}_{i-1}' axes is $N - 1$. The specific case for $N = 3$ and $i = 2$ is shown next in Fig. 8.31. The region $\overline{NH}_{2'}$ is computed using the vertices shown with the formula

for $\bar{\pi}_{2'}$. For $N = 3, i = 3$, the region $\overline{NH}_{3'}$ is identical to $\overline{NH}_{0'}$ but translated three units to the right. A more interesting example is in Fig. 8.30 for $N = 4, i = 3$; it is helpful for understanding general situation.

For small N , the vertices can be conveniently found by the traversal diagram. For the sake of completeness, it is useful to know for large k the sequence of vertices $A_{i',k}, B_{i',k} \in \partial \mathbf{B}$, and the c_r varying along some of the edges resulting from the i th shift are listed below.

$$\left\{ \begin{array}{l} A_1 = (c_1^-, \dots, c_N^-) \xrightarrow{c_i}, \\ A_{i',2} = (c_1^-, \dots, c_{i-1}^-, c_i^+, c_{i+1}^-, \dots, c_N^-), \\ A_{i',3} = (c_1^-, \dots, c_{i-1}^-, c_i^+, c_{i+1}^+, c_{i+2}^-, \dots, c_N^-), \\ \dots \\ A_{i',k} = (c_1^-, \dots, c_{i-1}^-, c_i^+, \dots, c_{i+k-1}^+, \dots, c_N^-), k < N - i + 1, \\ \dots \\ \xrightarrow{c_N} A_{i',N-i+1} = (c_1^-, \dots, c_{i-1}^-, c_i^+, \dots, c_{N-1}^+, c_N^+) \xrightarrow{c_{N+1}}, \\ A_{i',N-i} = (c_1^+, c_2^-, \dots, c_{i-1}^-, c_i^+, \dots, c_{N-1}^+, c_N^+) \xrightarrow{c_{N+2}}, \\ \dots \\ \xrightarrow{c_{i-2}} A_{i',N} = (c_1^+, c_2^+, \dots, c_{i-2}^+, c_{i-1}^-, c_i^+, \dots, c_N^+) \xrightarrow{c_{i-1}} \\ B_1 = (c_1^+, \dots, c_N^+) \xrightarrow{c_i} \\ B_{i',2} = (c_1^+, \dots, c_{i-1}^+, c_i^-, c_{i+1}^+, \dots, c_N^+) \xrightarrow{c_{i+1}} \\ \dots \\ \xrightarrow{c_N} B_{i',N-i+1} = (c_1^+, c_2^+, \dots, c_{i-1}^+, c_i^-, \dots, c_N^-) \xrightarrow{c_{N+1}} \\ \dots \\ \xrightarrow{c_{i-2}} B_{i',N} = (c_1^-, c_2^-, \dots, c_{i-2}^-, c_{i-1}^+, c_i^-, \dots, c_N^-) \xrightarrow{c_{i-1}} A_1 \xrightarrow{c_i} . \end{array} \right. \quad (8.54)$$

Corollary 8.3.11 (T. Matskewich) *Out of a total of 2^N vertices of $\partial \mathbf{B}$ $2N(N-1)$, those given in (8.54) are needed for the construction of the complete neighborhood \overline{NH} consisting of $N-1$ polygonal regions ($2N$ -gons).*

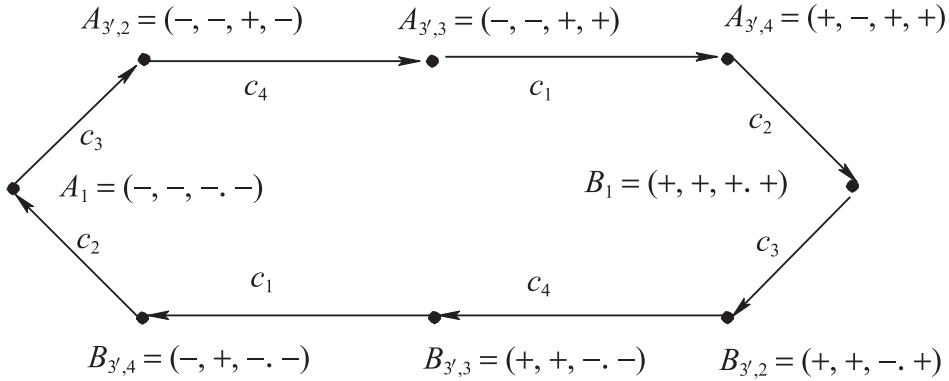


Figure 8.30. For $N = 4$ the shift is $i = 3$, i.e., $k \rightarrow k + 3 \pmod{4}$. This is the resulting path and vertices on $\partial \mathbf{B}$.

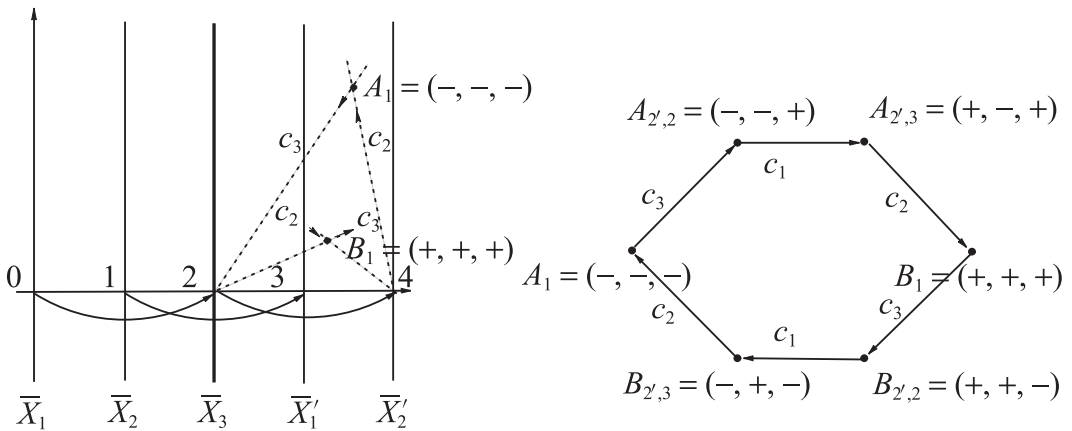


Figure 8.31. The shift by 2 units, $k \rightarrow k + 2 \pmod{3}$ and the resulting path on $\partial \mathbf{B}$.

Adding c_0

As was done for lines, there remains the task of enlarging the family of “close” hyperplanes by allowing c_0 to vary in the equation of π . The neighborhoods are now defined by

$$NH = \{\pi | \pi : c_1 x_1 + c_2 x_2 + \cdots + c_N x_N = c_0,$$

$$c_i \in [c_i^-, c_i^+], \quad i = 0, 1, 2, \dots, N\}. \quad (8.55)$$

When necessary, the notation NH_0 for a general c_0 is used to distinguish it from the previous neighborhoods with $c_0 = 1$. As for line neighborhoods, the y coordinates

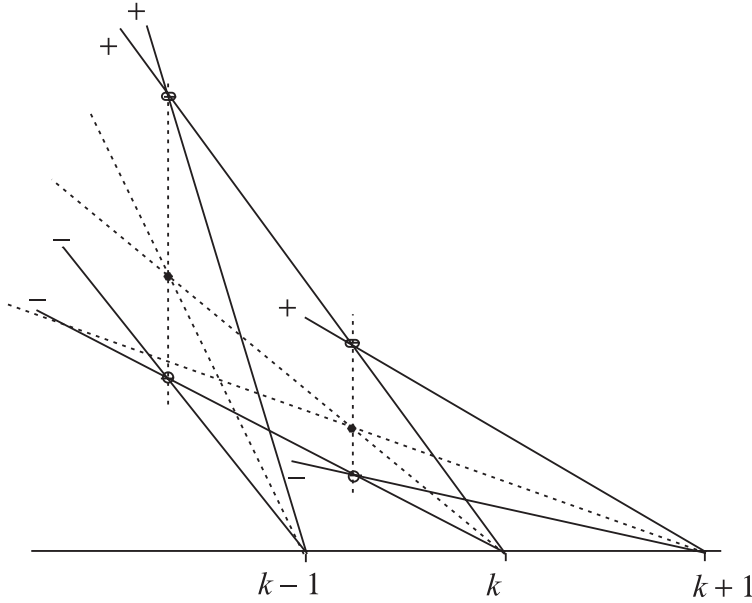


Figure 8.32. Construction of the vertices when $c_0 \neq 0$.

Instead of the original vertices shown in black circles, there are now two for c_0^+ , higher, and for c_0^- , both with the same x coordinate. The corresponding construction lines contributing to the new edges are shown.

of the vertices change according to c_0 , and two vertical edges are added, so that the polygons have $2(N+1)$ vertices and edges. The polygon still consists of the points $\bar{\pi}_Q$, which are now the function values of

$$f_N : \mathbb{R}^{N+1} \mapsto \mathbb{R}^2, \quad f_N(\mathbf{c}) = (x(\mathbf{c}), y(\mathbf{c})) = \left(\frac{\sum_{j=1}^N (j-1)c_j}{S}, \frac{c_0}{S} \right), \quad (8.56)$$

$$\mathbf{c} \in \mathbf{B}_0 = [c_0^-, c_0^+] \times [c_1^-, c_1^+] \times \cdots \times [c_N^-, c_N^+] \subset \mathbb{R}^{N+1},$$

an $(N+1)$ -dimensional box. The insertion of vertical edges poses a bit of a challenge for *uc* and *gh* regions, and the interesting details are left as exercises. Other than that, the addition of c_0 poses no difficulties. Repeating the previous decomposition

$$x = \frac{\sum_{j=1}^N (j-1)c_j}{S} = (k-1) + \frac{\sum_{j=1}^N (j-k)c_j}{S},$$

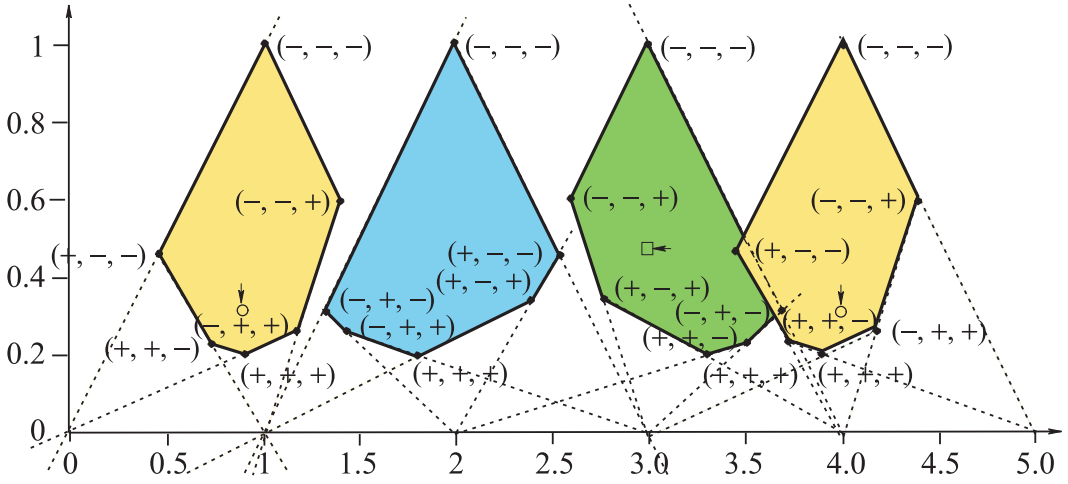


Figure 8.33. ♣ **FT-5** The four hexagonal regions.

Here the family of planes has $c_1 \in [1/3, 1.5]$, $c_2 \in [1/3, 2.5]$, $c_3 \in [1/3, 1]$. Note that the last region $\overline{NH}_{3'}$ is identical to the first $\overline{NH}_{0'}$ translated three units to the right. Note the overlap in the last two regions, which suggests that there may be planes π in this family with $c_3 = 0$ (when $\bar{\pi}_{2'} = \bar{\pi}_{3'}$), which is not possible from the definition of the c_k intervals. However, there cannot be points $\bar{\pi}_{2'} = \bar{\pi}_{3'} \in \overline{NH}_{2'} \cap \overline{NH}_{3'}$ with $\bar{\pi}_{0'} \in \overline{NH}_{0'}$ and $\bar{\pi}_{1'} \in \overline{NH}_{1'}$. For example, the two \bigcirc with vertical arrows are the locations of $\bar{\pi}_{0'}$, $\bar{\pi}_{3'}$ in $\overline{NH}_{0'}$ and $\overline{NH}_{3'}$ for $(-, +, -)$, while the \square with a horizontal arrow is the location of $(+, -, -)$ in $\overline{NH}_{2'}$.

the x coordinate of f_N , for $k = 1, 2, \dots, N$, is found in terms of the second coordinate, which is now $y = c_0/S$, just adding c_0 to the previous linear relation with the C^k defined as before:

$$\bar{P}^k : x = \frac{C^k}{c_0} y + (k - 1). \quad (8.57)$$

The net result in the construction of the polygons is that now for each k there are two lines

$$\bar{P}_{m,-}^k : x = \frac{C_m^k}{c_0^-} y + (k - 1), \quad \bar{P}_{m,+}^k : x = \frac{C_m^k}{c_0^+} y + (k - 1) \quad (8.58)$$

$$\bar{P}_{M,-}^k : x = \frac{C_M^k}{c_0^-} y + (k - 1), \quad \bar{P}_{M,+}^k : x = \frac{C_M^k}{c_0^+} y + (k - 1). \quad (8.59)$$

Figure 8.4 shows the situation for line neighborhoods, which basically remains the same here. As shown in Fig. 8.32, for each k there are now two points corresponding to c_0^+ and c_0^- with the same x coordinate obtained by the intersection of the

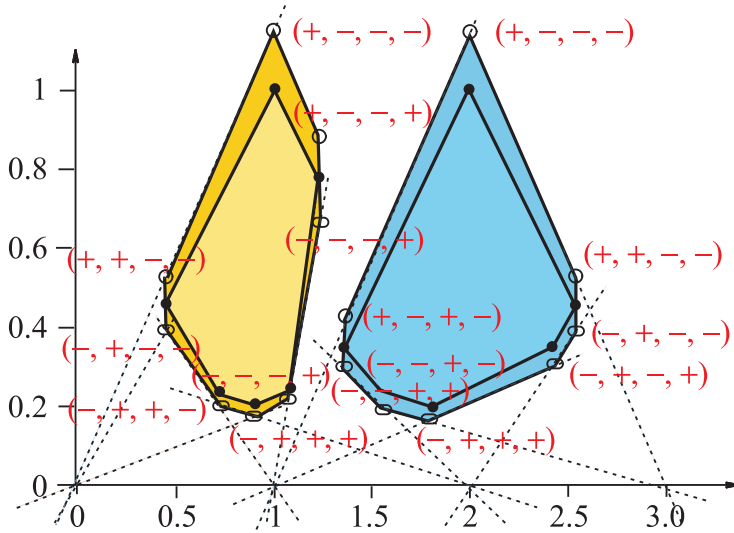


Figure 8.34. ♣ FT-6 The octagons $\Omega = \overline{NH}_0$ (left) and \overline{NH}_1 (right) for the family of planes with $c_0 \in [0.85, 1.15]$, $c_1 \in [1/3, 1.5]$, $c_2 \in [1/3, 2.5]$, $c_3 \in [1/3, 1]$. The superscript of c_0 is the first entry of the quadruples (\pm, \pm, \pm, \pm) designating the vertices. Vertices with c_0^+ are marked by \bigcirc and those for c_0^- with ellipses.

corresponding \bar{P} 's with \pm subscripts respectively. Though c_0 does not affect the type of \overline{NH} , which is determined by $S(\mathbf{c}) = \mathbf{0}$ if $0 \in [c_0^-, c_0^+]$, the x axis intersects the region if $c_0^- < 0$, requiring an amendment of Theorem 8.3.10. All this becomes clearer as we consider some examples.

Theorem 8.3.12. *For $c_0 \in [c_0^-, c_0^+]$, the region Ω is a $2(N + 1)$ -gon, which may have its leftmost or rightmost (or both) edges vertical and which is*

1. **bc** if $\pi_c \cap B = \emptyset$,
2. **uc** if $\pi_c \cap B = \{A_1\}$ or $\pi_c \cap B = \{B_1\}$,
3. **gh** if π_c intersects B at more than one edge, and has a vertical asymptote if in addition, π_c contains a vertex of B .

8.3.7 Some Examples

A few concrete examples illustrate some special cases in the process, showing that the constructions are basically easy. The relations

$$x(\bar{\pi}_{1'2'3}) = 6 - x(\bar{\pi}_{123}) - x(\bar{\pi}_{1'23}), \quad x(\bar{\pi}_{1'2'3'}) = 3 + x(\bar{\pi}_{123}), \quad (8.60)$$

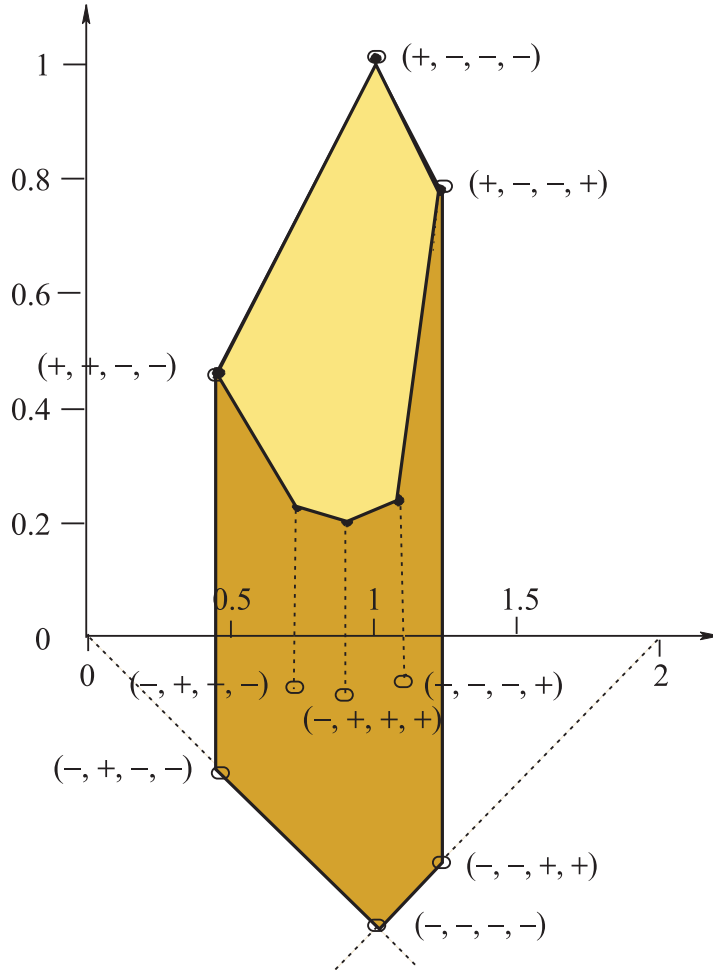


Figure 8.35. The octagon \overline{NH}_0' obtained from the hexagon in Fig. 8.26 with $c_0 \in [-0.5, 1]$. The x -axis is crossed, since $c_0 = 0$ is included in the range.

and their higher-dimensional generalizations given in Chapter 5 turn out to be very useful in the computations. We complete the previous $3D$ example in Fig. 8.26, constructing the hexagon $\overline{NH}_{2'}$ using the vertices for the shift $k \rightarrow k + 2 \bmod 3$ shown in Fig. 8.31. The fourth region $\overline{NH}_{3'}$, see (8.60), is simply a translation of \overline{NH}_0' three units to the right. Here we encounter an unexpected overlap between these two hexagons, suggesting that there may be planes π in this family with $\bar{\pi}_{2'} = \bar{\pi}_{3'} \Rightarrow c_3 = 0$, which is not possible, since here $c_3 \in [1/3, 1]$. For the fun of it, for a couple of coincident points on this overlap, their corresponding $\bar{\pi}_0', \bar{\pi}_1'$ are

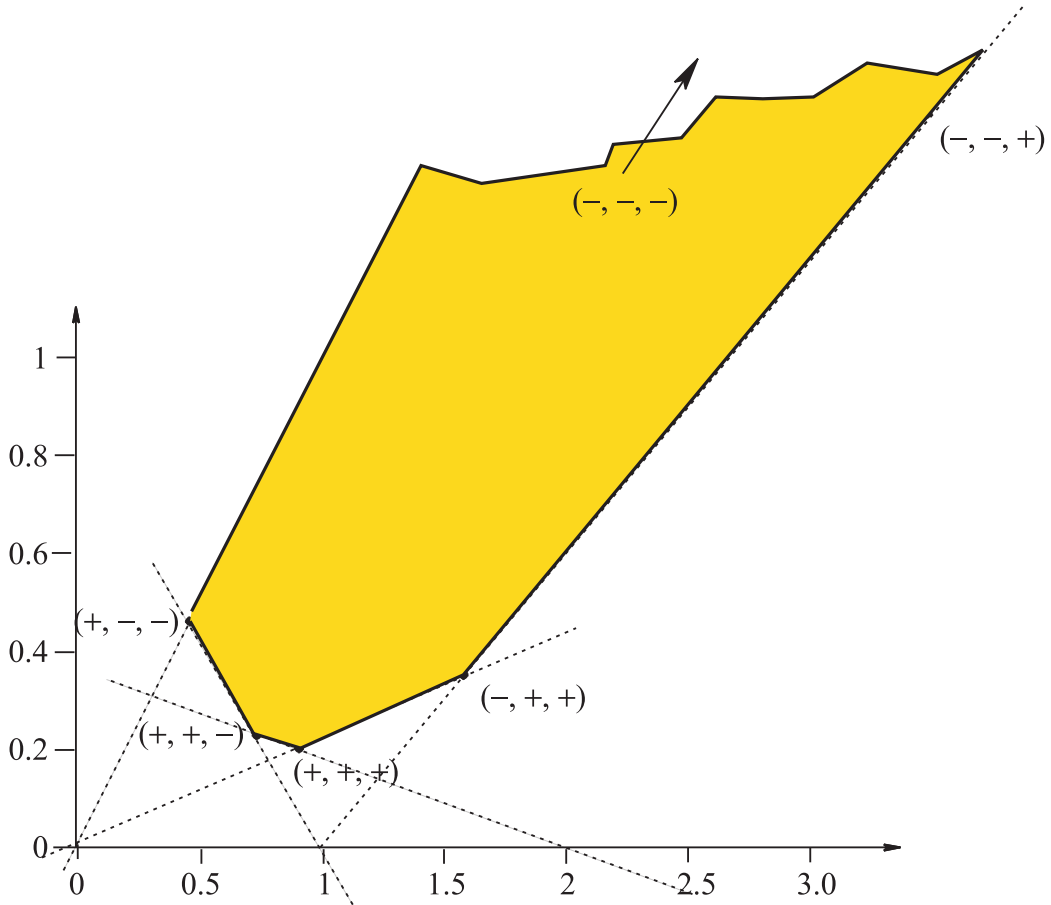


Figure 8.36. A uc in 3D for $c_1 \in [-2/3, 1.50]$, $c_2 \in [1/3, 2.50]$, $c_3 \in [1/3, 1.0]$.

Compare with the coefficient ranges of the hexagon(s) in Fig. 8.26 and note that they differ only for c_1^- . Here the sum $c_1^- + c_2^- + c_3^-$ is equal to 0.

found, and they differ, showing that these belong to two different planes. All this is detailed in the caption of Fig. 8.33.

Still working with Fig. 8.26 but now with variable $c_0 \in [0.85, 1.15]$, two octagons are produced with two vertical edges as shown in Fig. 8.34. For $c_0 < 0$ the resulting region crosses the x axis; Fig. 8.35. By the way, the appearance of vertical edges does not necessarily imply that c_0 is varying. For instance, in the family of planes in Fig. 8.26, just changing $c_1^- = 1$, which is also the value of c_3^+ , produces the hexagon \overline{NH}_0 , whose leftmost edge of \overline{NH}_0 , on $(-, -, +)$ and $(-, +, +)$, is vertical (see Exercise 9). Minimally varying this family of planes

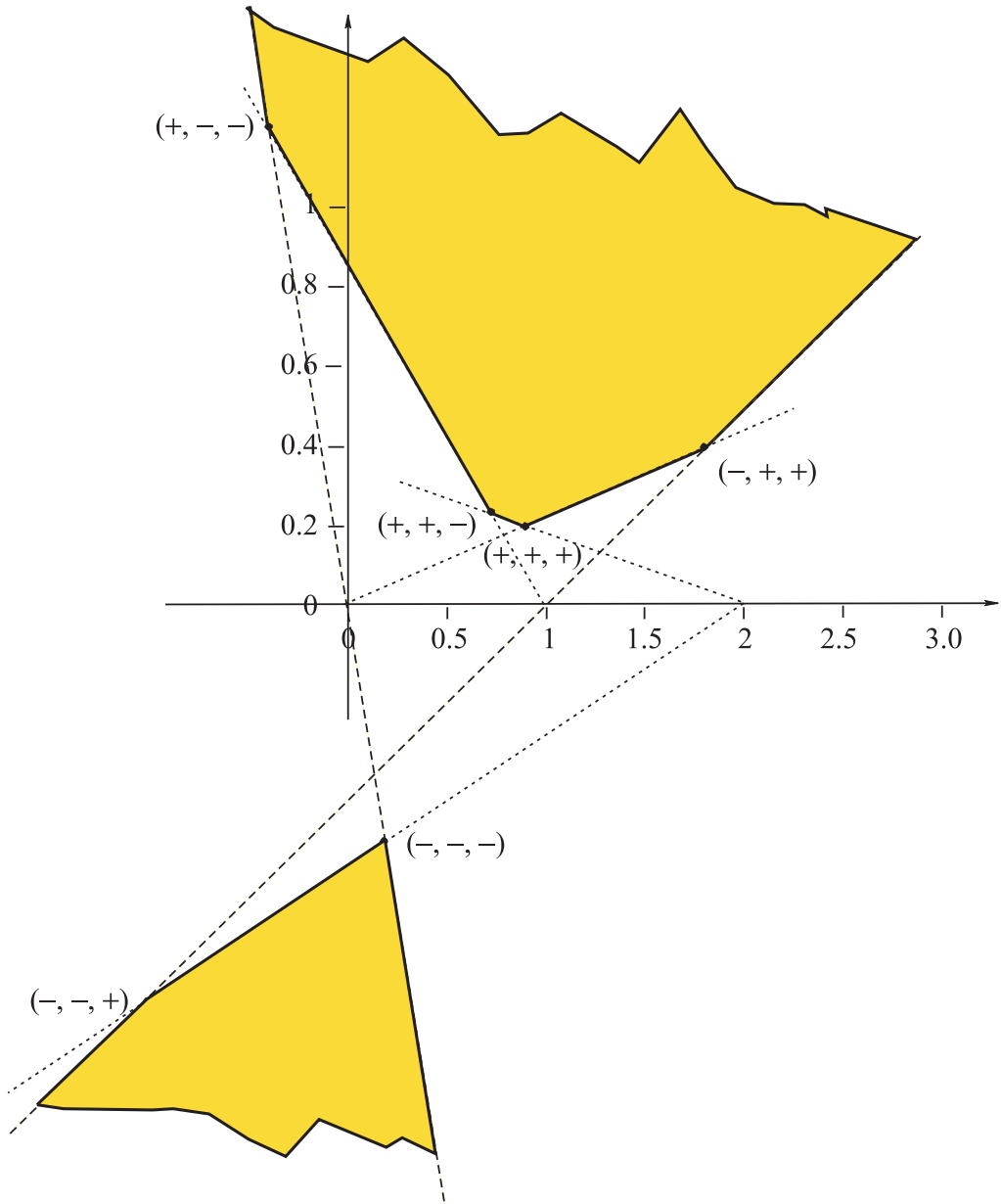


Figure 8.37. A uc in 3D for $c_1 \in [-1, 1.50]$, $c_2 \in [-1, 2.50]$, $c_3 \in [1/3, 1.0]$.

Compare with the coefficient ranges of the hexagon(s) in Fig. 8.26 and note that they differ only for c_1^-, c_2^- . Here the plane $\pi_c : c_1 + c_2 + c_3 = 0$ intersects two edges of the box \mathbf{B} with the vertices $(-, -, -)$, $(-, -, +)$ below π_c .

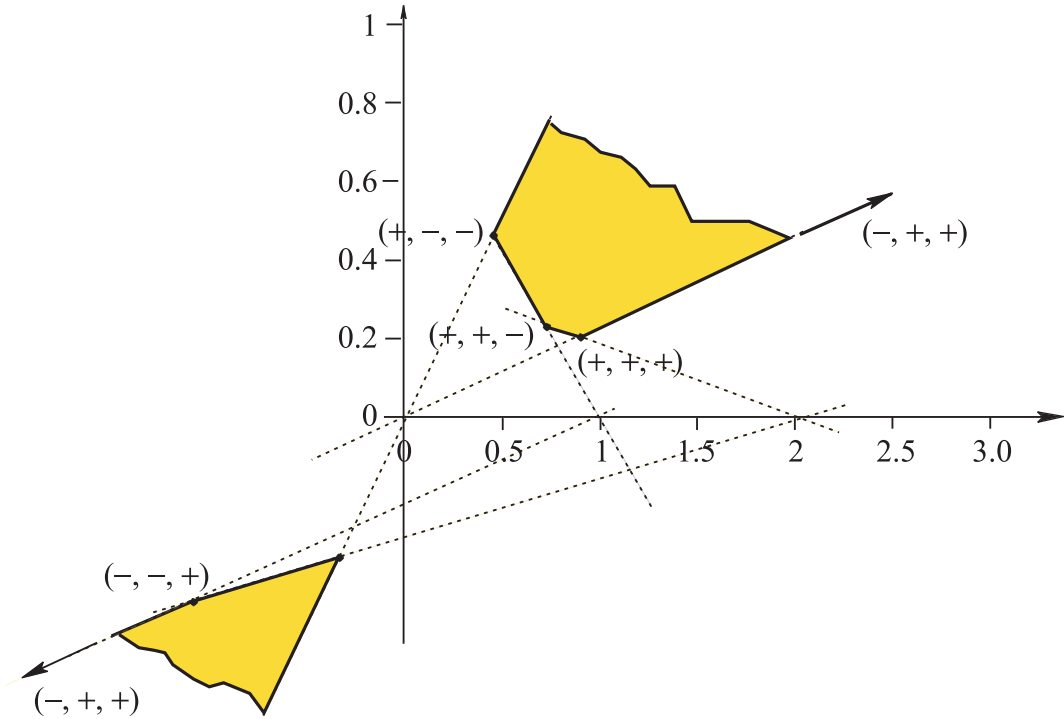


Figure 8.38. A gh in 3D for $c_1 \in [-3.5, 1.50]$, $c_2 \in [1/3, 2.50]$, $c_3 \in [1/3, 1.0]$.

Compare with the coefficient ranges of the hexagon(s) in Fig. 8.26 to note that they differ only in c_1^- . Here the plane $\pi_c : c_1 + c_2 + c_3 = 0$ contains the vertex $(-, +, +)$ of $\partial \mathbf{B}$, which becomes an ideal point of the gh .

shows how the regions evolve from bc to uc (Fig. 8.36), then gh (Fig. 8.37) and gh with an ideal point (Fig. 8.38). The ideal point of the uc is computed directly from the function (8.49). We then move on to 4D with the sequence of Figures 8.40, 8.41, 8.42, 8.43 having already pointed out that the ideal point of a uc is computed directly from the function f_N .

8.3.8 Matching Points and Navigation within \overline{NH}

The \overline{NH}_i contain all the points representing the Π family of planes given by (8.17). Though not all the vertices of $\partial \mathbf{B}$ are used in the construction of the $2N$ -vertex regions \overline{NH}_i , their images play an important role in delimiting which $N - 1$ points with the same y , one in each \overline{NH}_i , $i = 0, 2, \dots, N - 2$, represent a plane of the family Π . After all, there are linear relations, those defining the family Π , which introduce additional constraints connecting these points. The problem and its

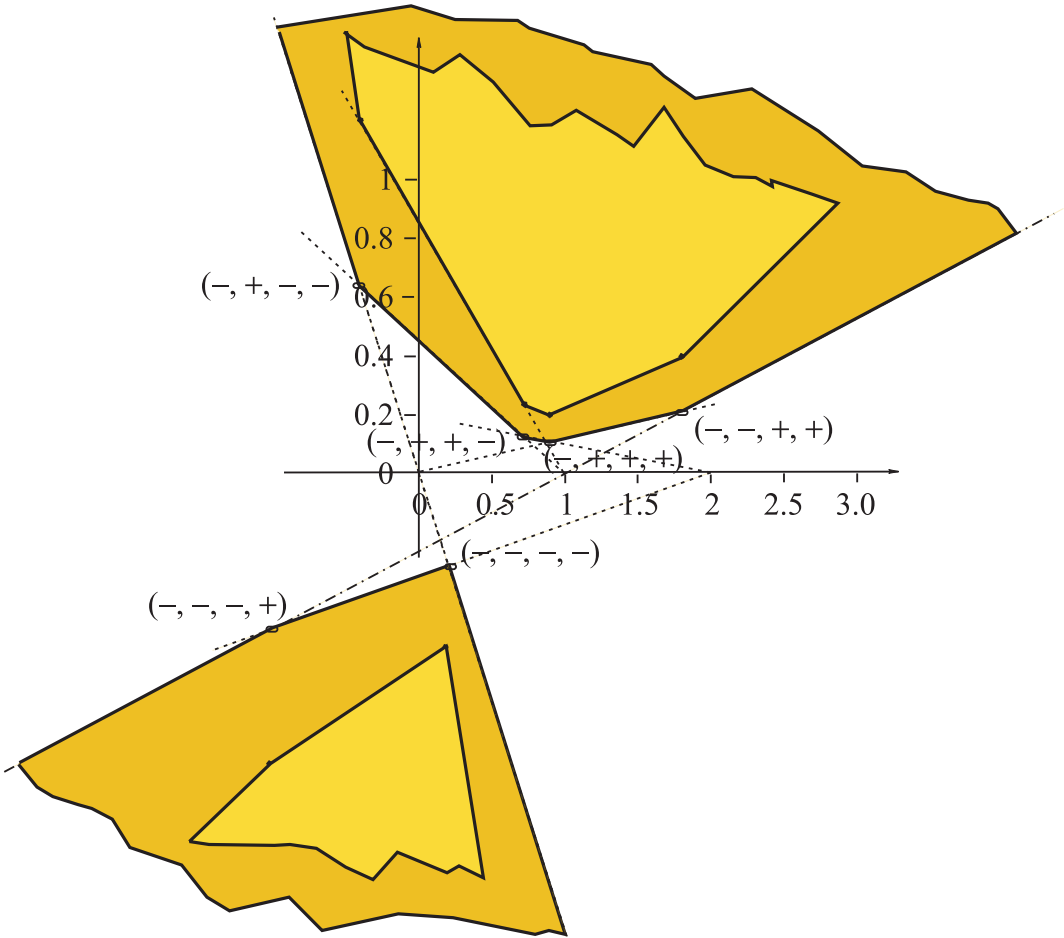


Figure 8.39. The gh in Fig. 8.37 with the range $c_0 \in [0.5, 1]$ added.

In the quintuplet $(., ., ., ., .)$, the first position is for the superscript of c_0^\pm . The previous vertices shown in Fig. 8.37, marked with \bullet , belong to the gh with $c_0^+ = 1$.

solution are illustrated using the previous family of 3D planes whose four hexagons are shown in Fig. 8.44. The images of all vertices of $\partial \mathbf{B}$ are shown, with each hexagon containing two additional interior points.

Two points with the same y , shown in Fig. 8.44, are in \overline{NH}_0 and \overline{NH}_1 , but they *cannot* represent a plane π in Π , since $\bar{\pi}_{2'} \notin \overline{NH}_3$. From the position of the points the represented plane π is found having coefficients $c_1 = 0.5$, $c_2 = 0.65$, $c_3 = 0.18$; hence $\pi \notin \Pi$, since $\bar{\pi}_{2'} \notin [1/3, 1]$. With this example and upon reflection, it is clear that *any* $N - 1$ points with the same y , one in each \overline{NH}_i ,

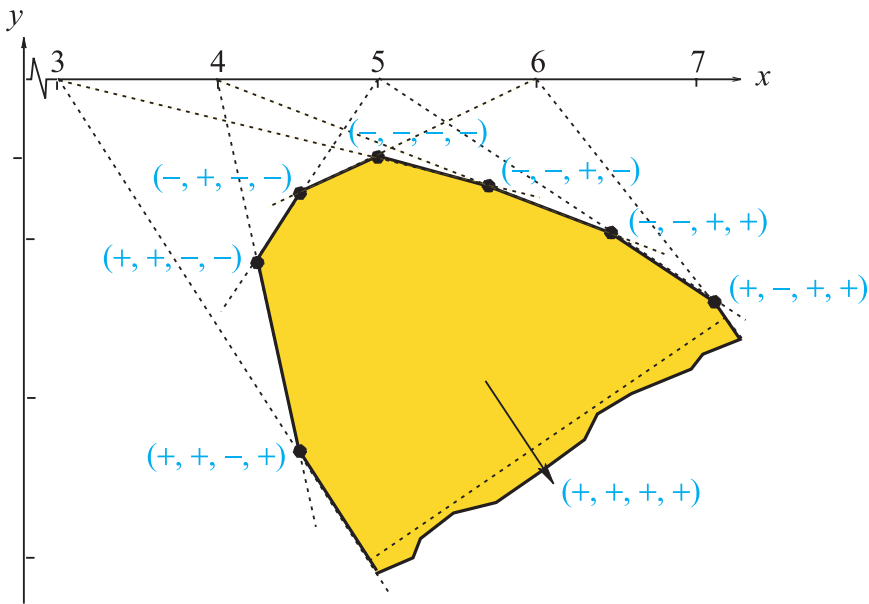


Figure 8.40. A \overline{NH}_3' uc in 4D constructed according to the traversal diagram in Fig. 8.30.

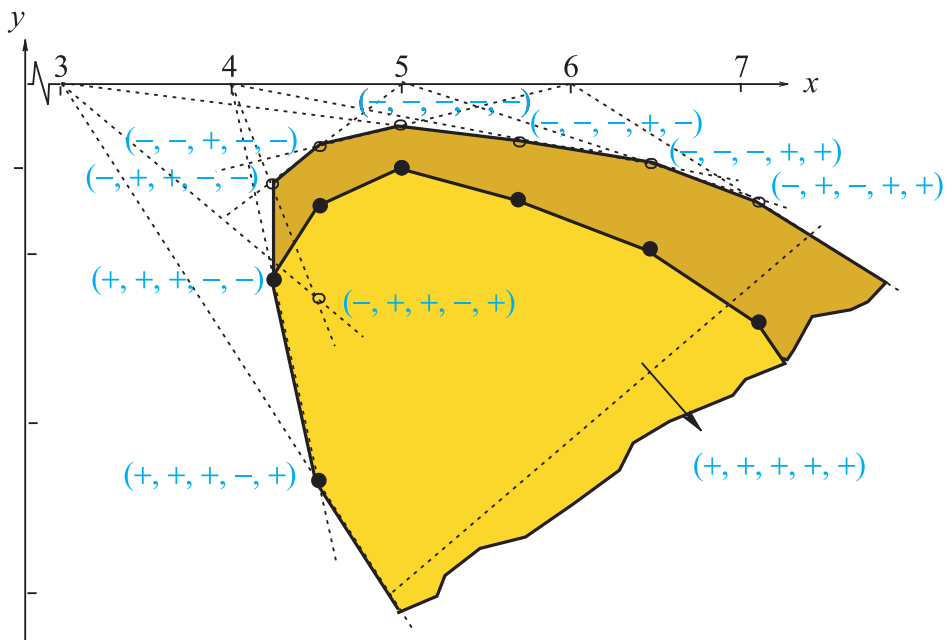


Figure 8.41. The previous $uc \overline{NH}_3'$ region in 4D with $c_0 \in [0.5, 1]$.

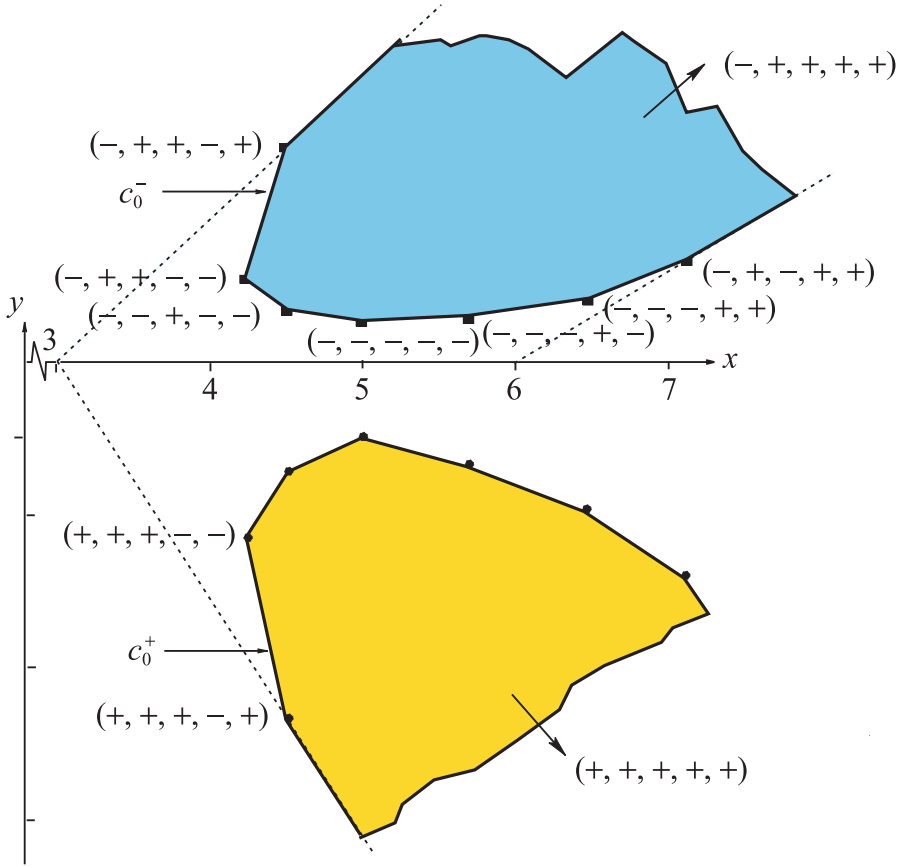


Figure 8.42. The two extreme regions for the previous $uc \overline{NH}_{3'}$ in 4D with $c_0 \in [-0.5, 1]$. Above is the $\overline{NH}_{3'}$ for $c_0 = -0.5$ and below for $c_0 = 1$.

$i = 0, 1, \dots, N-2$, do *not* necessarily represent a plane in the family Π . Exploring further, the same point $\pi_{0'}$ marked with \bullet on the edge $(-, -, -)$, $(+, -, -)$ of $\overline{NH}_{0'}$ together with a companion point $(\bar{\pi})_{1'}$ with the same y coordinate on the edge $(-, -, -)$, $(+, -, -)$ of $\overline{NH}_{1'}$ (not shown) represent the plane with coefficients $c_1 = 0.44$, $c_2 = 0.76$ and $c_3 = 0.13$, which is also not a plane in Π . Next, consider the pair $\bar{\pi}_{0'}$ marked with \square and the previous $\bar{\pi}_{1'} \in \overline{NH}_{1'}$, the point marked with both \square and \bullet , and $\bar{\pi}_{2'} \in \overline{NH}_{2'}$, marked with a \square . These represent the plane π in Π having coefficients $c_1 = 0.33$, $c_2 = 0.55$, $c_3 = 0.45$.

So what is going on? The dotted line joining the vertices $(-, -, -)$, $(-, +, -)$ together with the edges on $(-, -, -)$, $(-, -, +)$ and on $(-, -, +)$, $(-, +, -)$

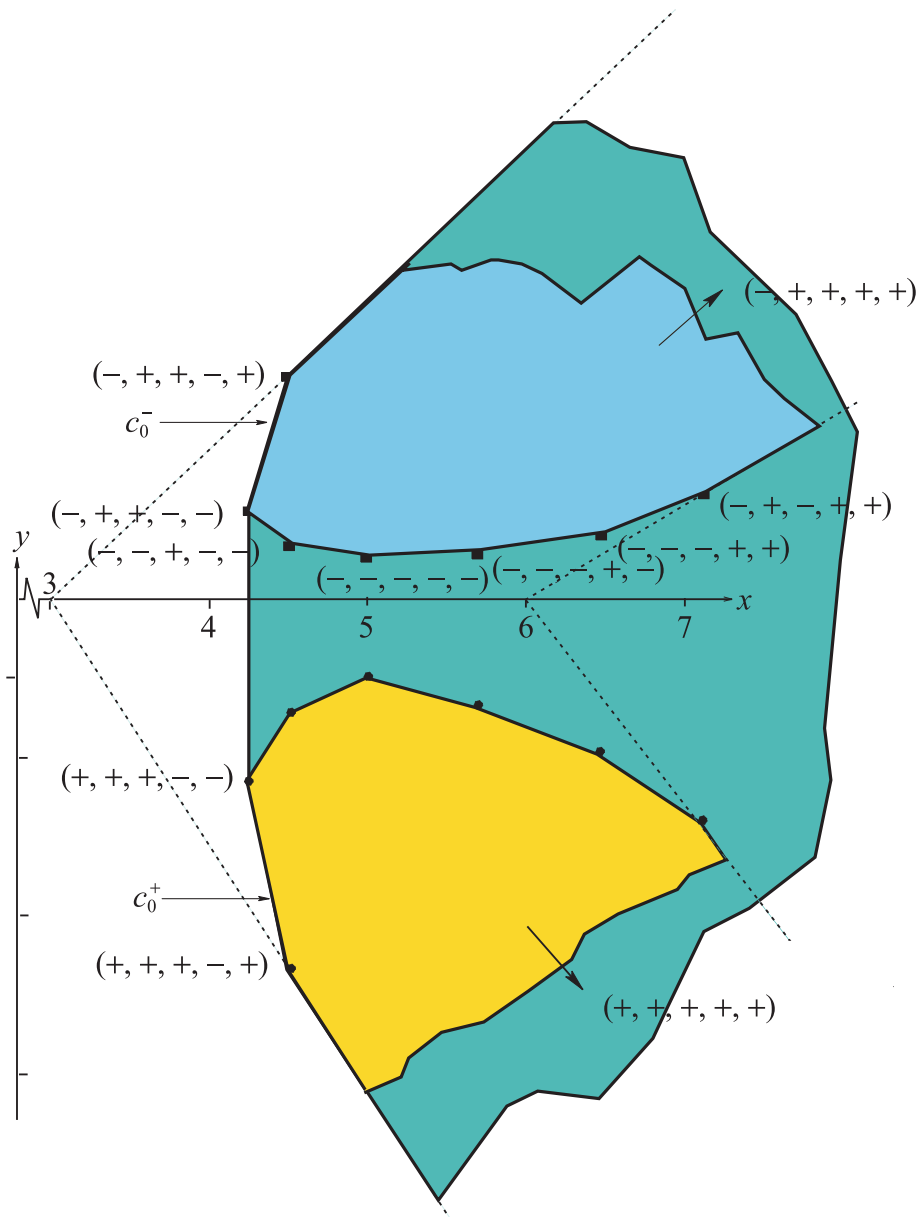


Figure 8.43. The full $uc \overline{NH}_{3'}$ in 4D with $c_0 \in [-0.5, 1]$. Note the vertical edge on the left.

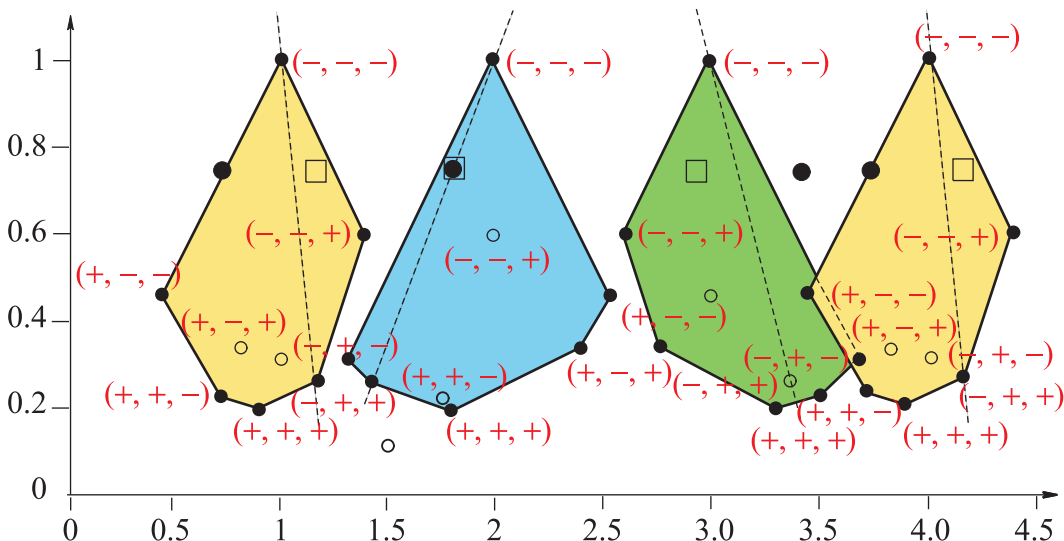


Figure 8.44. The four hexagons previously seen in Fig. 8.33.

For a family of planes Π with $c_1 \in [1/3, 1.5]$, $c_2 \in [1/3, 2.5]$, $c_3 \in [1/3, 1]$ now showing the images of *all* vertices of $\partial\mathbf{B}$, two being interior points. Though the first two points with the same y , marked by \bullet , are in \overline{NH}_1 and \overline{NH}_2 respectively, they do *not* represent a plane in Π , since $\bar{\pi}_2 \notin \overline{NH}_3$.

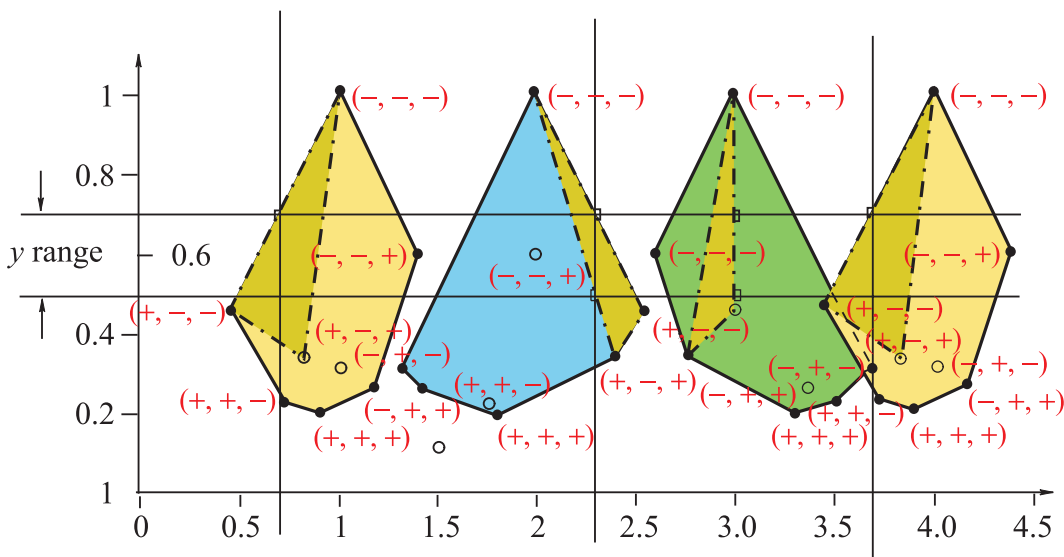


Figure 8.45. Finding interior points.

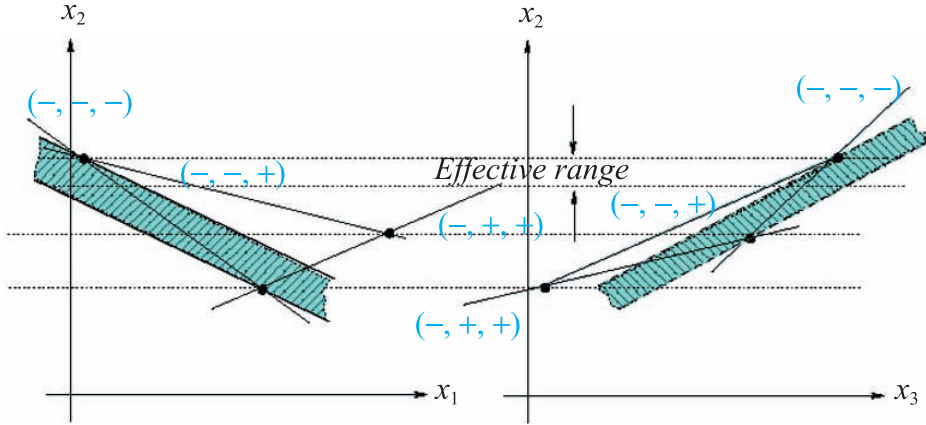


Figure 8.46. Triangulation in orthogonal coordinates.

Two sections of the family of planes Π perpendicular to the x_3 and x_1 axes. Focus on the intersection point P of the lines $(-, -, -)$, $(-, -, +)$ shown in the x_1x_2 and x_2x_3 planes. Choose the x coordinates of $\bar{\pi}_0, \bar{\pi}_1$ to fix a $\bar{\pi} \in \Pi$ and on P . The slopes in the x_1x_2 and x_2x_3 planes are determined from the x -coordinates of $\bar{\pi}_0, \bar{\pi}_1$ respectively. Generate this family of planes parallel to π and in Π by allowing x_2 to vary until it hits the constraint; here the lowest point. The range of x_2 differs in the two sections, and the effective range is the *minimum*. This corresponds to the triangulation procedure shown above in $\|\cdot\|$ -coords.

form a triangle in \overline{NH}_0' , and also in other \overline{NH}_i' , as shown in Fig. 8.45. Choose a point $\bar{\pi}_0'$ on the edge $(-, -, -)$, $(-, -, +)$ on \overline{NH}_0' , marked with \square , and on the same edge \overline{NH}_1' its companion point $\bar{\pi}_1'$ with the same y . Clearly, this π is in Π with coefficients $c_2 = c_2^- = c_3 = c_3^- = 1/3$. Any pair of points, one on each of the vertical lines, through these two points represents a plane π' parallel to π and hence with coefficients $c'_2 = c'_3$ by the previous construction for parallel planes. The range of y for which $\pi' \in \Pi$ is determined by the *highest* intersection of the verticals with an edge of the triangle, in this case with the triangle \overline{NH}_1' , as shown. The corresponding situation in Cartesian coordinates is illustrated in Fig. 8.46. To find further interior points with the same x coordinates, variation of the y range may be possible in some cases (see Fig. 8.47) but not others (Fig. 8.48) by further triangulation. It is best to study these pictures to understand what is happening. There is an important caveat to avoid vertical edges in the triangulation, and when one is encountered, then use its image in another \overline{NH}_i' , as was done in Fig. 8.47. The remaining details are left for the exercises and future generations. Suffice it to say that this procedure generalizes directly to \mathbb{R}^N .

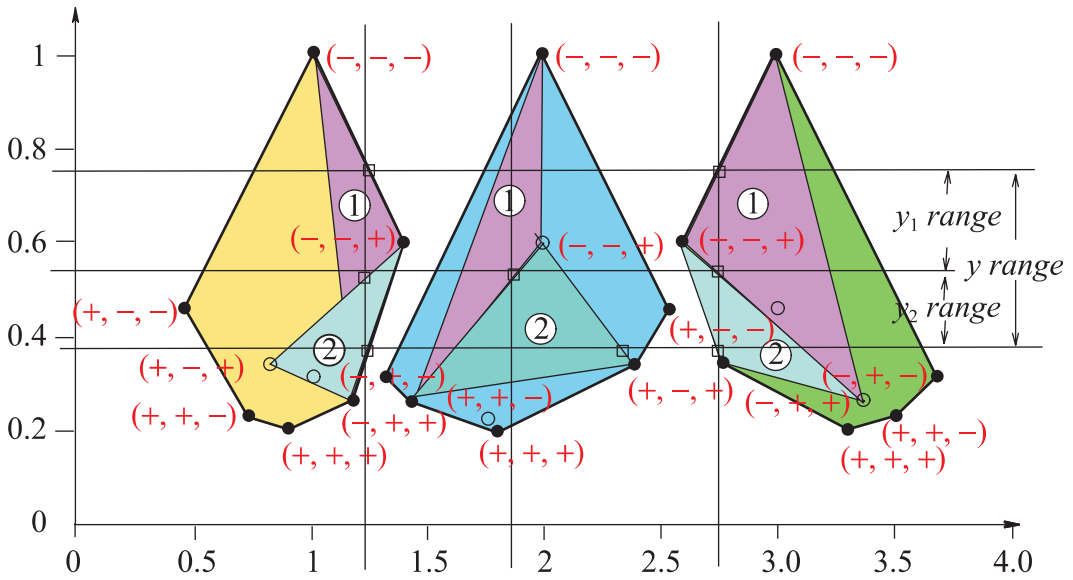


Figure 8.47. Continuing the interior point construction from triangle 1 to triangle 2.

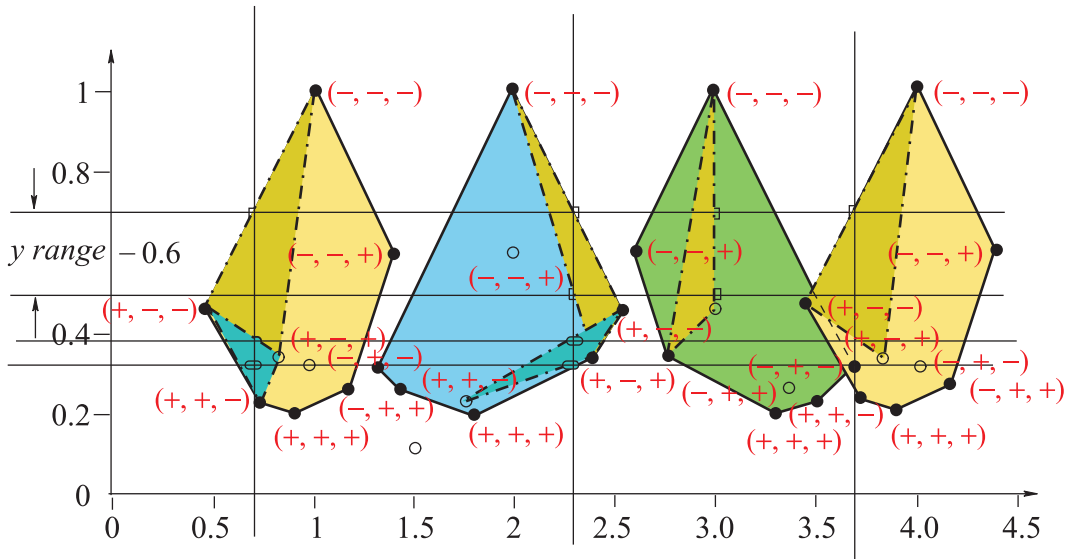


Figure 8.48. Continuation along the vertical by extending the triangulation. Here it is not possible due to gap between the two triangles in NH_2 .

Exercises

1. Investigate the neighborhood family \mathcal{N}_π of hyperplanes given by (8.17). Do they form a topology with respect to the criteria set in Section 8.1? This is a good topic for a *course project*.
2. Find the relations corresponding to (8.10), (8.11), (8.12) but for the indexed points $\bar{\pi}_{1234}, \bar{\pi}_{1'234}, \bar{\pi}_{1'2'34}, \bar{\pi}_{1'2'3'4}, \bar{\pi}_{1'2'3'4'}$ of a hyperplane $\pi \subset \mathbb{R}^4$.
3. Prove analytically the equivalences translation \leftrightarrow rotation, rotation \leftrightarrow translation indicated in Fig. 8.9.
4. Draw the equivalent of Fig. 8.21 in the space of coefficients \mathbb{R}^N using the box B in Fig. 8.20.
5. Draw the construction of Ω when π_c intersects B at a vertex A_s and an edge b_r . Hint: \bar{A}_k is a vertical asymptote of Ω .
6. Construct and draw the *uc*, *gh* cases for the *line* neighborhoods Ω in \mathbb{R}^N . Give specific examples for $N = 3, 4$.
7. For the neighborhoods given by (8.55) with $c_0 \neq 1$, carry out the analysis of the resulting shape of the regions Ω , and construct and draw examples of the possible options. State the result formally as a corollary to Theorem 8.3.10. This together with a software implementation where the user specifies the intervals $[c_k^-, c_k^+]$, and the region Ω is drawn, is an excellent topic for a *course project*.
8. Give examples of polygonal regions with $2N$ vertices/edges and prove that they can *not* be the output of the “*scissors algorithm*.”
9. Draw some $\bar{N}H_0$ for $2N$ -gons *uc* and *gh* with vertical edges and c_0 .
10. Write an algorithm for finding interior points in \mathbb{R}^N for $N = 3, 4$. This together with a software implementation is an excellent topic for a *course project*.

This page intentionally left blank

Hypersurfaces in \mathbb{R}^N

9.1 *FT-1* Preliminaries

Early in the development of \parallel -coords, the representation of surfaces was attempted by simply plotting the surface's points as polygonal lines. Except for very special surfaces (i.e., spheres, ellipsoids), this turned out to be even more hopeless than what happened with line-curves. The next improvement was the surface representation by the *envelope* of these polygonal lines. This is akin to using the projections in \parallel -coords of some surface sections. Since projections do not preserve information, this approach also does not work, though it is useful (but not accurate) at least for ellipsoids ([98], [81]), and for this reason it is discussed later on. However, it is a mess (worse than cusps and other special points in line-curves) for nonconvex regions. The way out turns out to be the natural continuation of curve images as treated in Chapter 7.

We start by revisiting Fig. 9.1, which was shown earlier, in Chapter 3. The top shows the hypercube representation and serves as an excellent paradigm for what is to follow. In (a) is a square in Cartesian and \parallel -coords. Note the role of duality where a vertex, say A , is mapped into a line \bar{A} , and conversely, an edge, say AB , is mapped into a point \bar{AB} . Proceeding, in (b) the image of the 3-D cube is *twice* the pattern of the square, where the edges can be identified by two points and faces by two vertical lines and a polygonal line. This is a good example, showing where the general representation of lines in Section 4.1.4 [106] is needed for the edges with two of the three variables constant (in effect, this simplifies matters). The representation of the cube's faces stems from the vertical-line representation of hyperplanes discussed in Chapter 5. The pattern's repetition pleasingly reveals the cube's symmetry and its relation to the square. Whereas we do not know how to show a hypercube in Cartesian coordinates, it is no problem doing so in

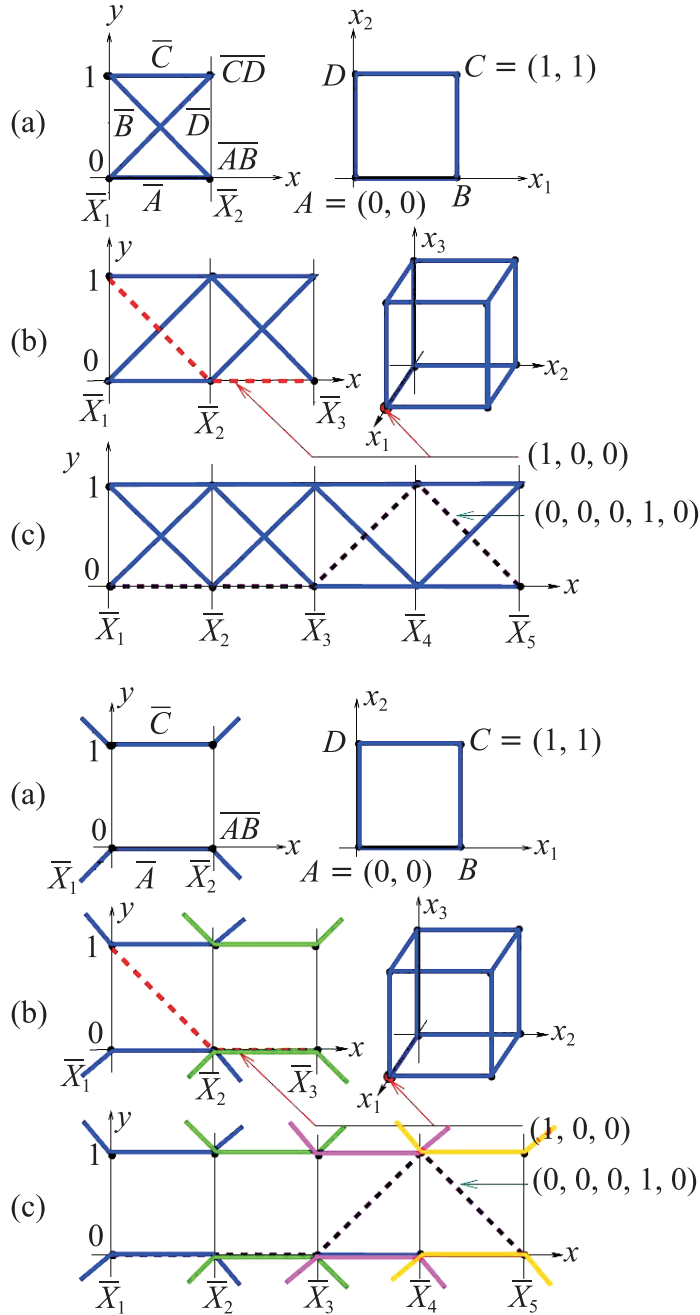


Figure 9.1. (a) Square (b) Cube in \mathbb{R}^3 (c) Hypercube in \mathbb{R}^5 – all edges have unit length. (top) Showing the vertices and (bottom) representation based on the tangent planes as developed in this chapter.

$\|\text{-coords}$. In (c) is the representation of a hypercube in \mathbb{R}^5 consisting of four times the square's pattern. Again all vertices, edges, and faces can be determined from the picture. Interior points are represented by polygonal lines *within* the pattern. Polygonal lines that include a boundary segment represent boundary points. The bottom part of Fig. 9.1 depicts the new representation developed in this chapter. This suggests how the representation of higher-dimensional versions of a surface may be obtained from its 3-D representation. All this generalizes nicely to the representation of polytopes in \mathbb{R}^N [23]. A rotated or translated hypercube can still be easily represented and recognized (exercises). The complete description of polyhedral objects is conveniently given in terms of *incidence matrices*, which are tables listing the vertices, the edges, and faces of all orders that they are on (i.e., their incidence) (see [146, pp. 139–140]).

To lay the groundwork, recall the recursive notation for the axis spacing (considered as an N -tuple) developed in Chapter 5 on planes. For the initial N -axis system $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N$, we write

$$\mathbf{d}_N^0 = \overbrace{(0, 1, 2, \dots, i-1, \dots, N-1)}^i.$$

This states that the first axis \bar{X}_1 is placed at $x = 0$, \bar{X}_2 is placed at $x = 1, \dots$ until \bar{X}_N , which is at $x = N - 1$. After translating the \bar{X}_1 axis one unit to the right of \bar{X}_N , renaming it $\bar{X}_{1'}$, the axis spacing N -tuple is

$$\mathbf{d}_N^1 = \mathbf{d}_N^0 + (N, 0, \dots, 0) = \overbrace{(N, 1, 2, \dots, i-1, \dots, N-1)}^i.$$

And after i such successive unit translations with the \bar{X}_i axis in position $\bar{X}_{i'}$, the axis spacing for $i = 0, \dots, N, k = 1, \dots, N$ is given by

$$\begin{aligned} \mathbf{d}_N^i &= \mathbf{d}_N^0 + \overbrace{(N, \dots, N, 0, \dots, 0)}^i = \overbrace{(N, N+1, \dots, N+i-1, i, \dots, N-1)}^i \\ &= (d_{ik}). \end{aligned} \tag{9.1}$$

To clarify the indices, i is the number of axis translations and k is the position (component) within the vector \mathbf{d}_N^i . Using the step function

$$S_i(k) = \begin{cases} 1 & i \geq k, \\ 0 & i < k, \end{cases} \tag{9.2}$$

the axis spacing after the i th translation can be conveniently written as

$$\mathbf{d}_N^i = (d_{ik}) = (k - 1 + NS_i(k)). \tag{9.3}$$

When the dimensionality is clear from the context, the subscript N can be omitted. For a flat π^P expressed in terms of the \mathbf{d}_N^i spacing, the points $\bar{\pi}_{1', \dots, i', i+1, \dots, N}^P$ of its representation are denoted compactly by $\bar{\pi}_{i'}^P$. Consistent with this notation, $\pi^P = \pi_{0'}^P$ for π^P described in terms of the axis spacing \mathbf{d}_N^0 . So the common spacings are $\mathbf{d}_3^0 = (0, 1, 2)$, $\mathbf{d}_3^1 = (3, 1, 2)$, and so on.

It is useful to recast the representation of planes in vector form. For

$$\pi : c_1 x_1 + c_2 x_2 + c_3 x_3 = c_0, \quad (9.4)$$

the coefficient vector $\mathbf{c} = (c_1, c_2, c_3)$, and $\mathbf{u} = (1, 1, 1)$, the representing points of π given in inner-product (denoted by “ \cdot ”) form are:

$$\bar{\pi} = (\mathbf{c} \cdot \mathbf{d}_3^i, c_0, \mathbf{c} \cdot \mathbf{u}) = (\mathbf{c} \cdot \mathbf{d}_3^i, c_0, c_1 + c_2 + c_3). \quad (9.5)$$

The first coordinates of $\bar{\pi}$, with the axis spacing given by (9.3), are

$$\left\{ \begin{array}{l} \mathbf{c} \cdot \mathbf{d}^0 = \mathbf{c} \cdot (0, 1, 2) = c_2 + 2c_3, \\ \mathbf{c} \cdot \mathbf{d}^1 = \mathbf{c} \cdot (3, 1, 2) = 3c_1 + c_2 + 2c_3, \\ \mathbf{c} \cdot \mathbf{d}^2 = \mathbf{c} \cdot (3, 4, 2) = 3c_1 + 4c_2 + 2c_3, \\ \mathbf{c} \cdot \mathbf{d}^3 = \mathbf{c} \cdot (3, 4, 5) = 3c_1 + 4c_2 + 5c_3. \end{array} \right. \quad (9.6)$$

For $x_i = \mathbf{c} \cdot \mathbf{d}^i / S$ with $S = \sum_{i=0}^{N-1} c_i$, The relations

$$x_2 = 6 - [x_0 + x_1], \quad x_3 = 3 + x_0,$$

and in general for \mathbb{R}^N with $x_i = \mathbf{c} \cdot \mathbf{d}_N^i / S$ and $S = \sum_{i=1}^N c_i$,

$$x_{(N-1)} = N(N-1) - \sum_{i=0}^{N-2} x_i, \quad x_N = N + x_0,$$

provide the connection between the $N-1$ independent representing points and the remaining two.

Exercises

1. Show the 5-D hypercube in \parallel -coords after a
 - (a) translation,
 - (b) rotation, and
 - (c) both translated and rotated.
2. Construct *incidence matrices* for the 3-D and 5-D hypercube shown in Fig. 9.1.
3. Construct the representation in \parallel -coords and incidence matrices for an object (e.g., a pyramid) of your choice in 3-D and 5-D; see also [23].

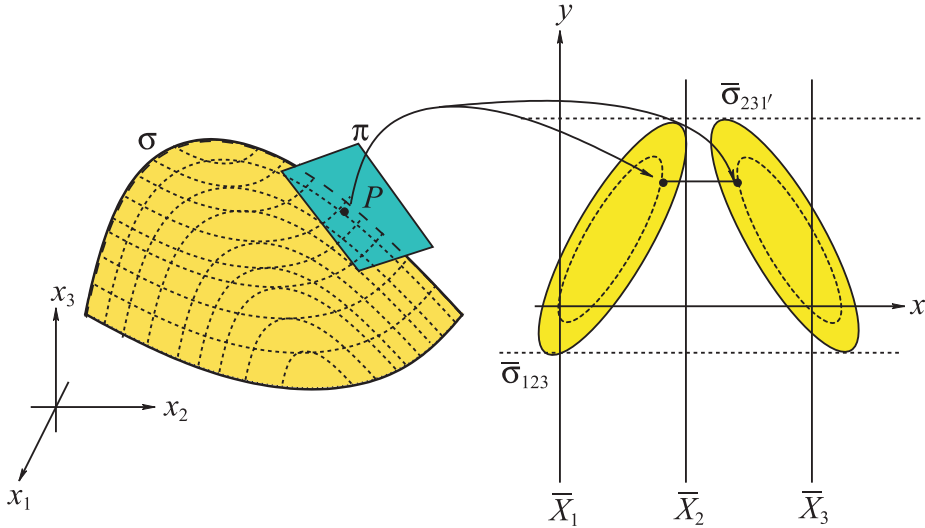


Figure 9.2. A surface $\sigma \in \mathcal{E}$ is represented by two linked planar regions $\bar{\sigma}_{123}, \bar{\sigma}_{231}'$. They consist of the pairs of points representing all its tangent planes.

9.2 Formulation

We now pursue the representation of smooth surfaces in \mathbb{R}^N and in particular the family \mathcal{E} of surfaces⁴⁰ that are the envelopes of their tangent planes. The discussion is confined to \mathbb{R}^3 when the higher-dimensional generalizations are clear. Occasionally, as for curves, we use terminology and results that can be found in the references on differential geometry. Also recommended is the beautiful and comprehensive website on surfaces [130]. The basic idea for the representation of a surface $\sigma \in \mathcal{E}$ is illustrated in Fig. 9.2. The tangent plane π at each point $P \in \sigma$ is mapped to the two planar points $\bar{\pi}_{123}, \bar{\pi}_{231}' = \bar{\pi}_{1'23}$. Collecting the points by their indices for all the tangent planes yields two planar regions $\bar{\sigma}_{123}(x, y), \bar{\sigma}_{1'23}(x, y)$, one each for the index triples. The regions must be *linked* by a *matching algorithm* that selects the pairs of points (one from each region) representing valid tangent planes of σ . The two linked regions form the representation $\bar{\sigma}$ of the surface. An example of the linking was already seen in Chapter 8 between the polygonal regions representing families of *proximate* planes. The intent is to *reconstruct* the surface from its representation. The manner and extent to which this is possible, i.e., when the $\|\cdot\|$ -coords mapping is invertible, are among the issues studied in this chapter.

⁴⁰Not to be confused with the family of superplanes also denoted by \mathcal{E} .

Formally, a surface $\sigma \in \mathcal{E}$ given by

$$\sigma : F(\mathbf{x}) = 0, \quad \mathbf{x} = (x_1, x_2, x_3), \quad (9.7)$$

is represented in $\|\text{-coords}$ by

$$\sigma \mapsto \bar{\sigma} = (\bar{\sigma}_{123}(x, y), \quad \bar{\sigma}_{1'23}(x, y)) \subset \mathbb{P}^2 \times \mathbb{P}^2, \quad (9.8)$$

where use of the projective plane \mathbb{P}^2 allows for the presence of ideal points. In the notation, the *link* between the two regions is indicated by placing them within the $(\ , \)$, which are omitted when the discussion pertains to just the region(s). The functions used are assumed to be continuous in all their variables together with such of its derivatives as are involved in the discussion.

The gradient vector of F , $\nabla F = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \frac{\partial F}{\partial x_3} \right) \big|_P$, at the point P is normal to the surface σ at $P \Rightarrow$ the tangent plane π of σ at the point $P_0(\mathbf{x}^0)$, $(\mathbf{x}^0) = (x_1^0, x_2^0, x_3^0)$ is given by

$$\pi : \nabla F \cdot (\mathbf{x} - \mathbf{x}^0) = \sum_{i=1}^3 (x_i - x_i^0) \frac{\partial F}{\partial x_i} (x_1^0, x_2^0, x_3^0) = 0. \quad (9.9)$$

Therefore the representing points of π are

$$\bar{\pi}_i(s, t) = (\nabla F \cdot \mathbf{d}^i, \nabla F \cdot (\mathbf{x}^0), \nabla F \cdot \mathbf{u}), \quad i = 0, 1. \quad (9.10)$$

For our purposes it is often preferable to describe a surface in terms of two parameters as

$$\sigma : F(s, t) = F(\mathbf{x}) = 0, \quad \mathbf{x} = \mathbf{x}(\mathbf{s}, \mathbf{t}), \quad s \in I_s, t \in I_t, \quad (9.11)$$

where I_s, I_t are intervals of \mathbb{R} . The equivalent description for hypersurfaces in \mathbb{R}^N requires $N - 1$ parameters. This form is due to the great mathematician Gauss. The vector $\mathbf{x}_s(\mathbf{s}, \mathbf{k}_2)$ is tangent to the surface curve $t = k_2$, and $\mathbf{x}_t(\mathbf{k}_1, \mathbf{t})$ is tangent to the surface curve $s = k_1$, the $k_i = \text{constant}$, and the subscripts denote partial derivatives with respect to the indicated variables. Therefore, the vector $\mathbf{x}_s \times \mathbf{x}_t$ is normal to the point $P = F(\mathbf{x}^0) \in \sigma$, where $\mathbf{x}^0 = \mathbf{x}(\mathbf{k}_1, \mathbf{k}_2)$, and the tangent plane π at $P_0(\mathbf{x}^0)$ is conveniently given by

$$\pi(s, t) : (\mathbf{x} - \mathbf{x}^0) \cdot \mathbf{x}_s \times \mathbf{x}_t = \begin{vmatrix} x_1 - x_1^0 & x_2 - x_2^0 & x_3 - x_3^0 \\ x_{1t} & x_{2t} & x_{3t} \\ x_{1s} & x_{2s} & x_{3s} \end{vmatrix} = 0. \quad (9.12)$$

Alternatively, $\sigma \in \mathcal{E}$ can be considered as the envelope of the 2-parameter family $\pi(s, t)$ of its tangent planes, each contributing a pair of points, one each for

the planar regions $\bar{\sigma}_{123}, \bar{\sigma}_{1'23}$, constituting the image $\bar{\sigma}$. The representing points of the tangent planes are equivalently given by

$$\bar{\pi}_{i'}(s, t) = (\hat{\mathbf{n}} \cdot \mathbf{d}^{\mathbf{i}}, \hat{\mathbf{n}} \cdot (\mathbf{x}^0), \hat{\mathbf{n}} \cdot \mathbf{u}), \quad i = 0, 1, \quad (9.13)$$

where

$$\hat{\mathbf{n}} = \frac{\mathbf{x}_s \times \mathbf{x}_t}{|\mathbf{x}_s \times \mathbf{x}_t|}$$

is the unit normal vector to σ at the point P_0 . Stated explicitly the point \mapsto pair-of-points mapping, from either (9.10) or (9.12), is

$$\mathbf{x} \in \sigma \mapsto \pi \mapsto (\bar{\pi}_{0'}, \bar{\pi}_{1'}) = (\bar{\pi}_{123}, \bar{\pi}_{1'23}) = ((x, y), (x', y)), \quad (9.14)$$

with a slight change in notation, $x = x_{0'}$ and $x' = x_{1'}$ for the x -coordinates of $\bar{\pi}_{123} = \bar{\pi}_{0'}$, $\bar{\pi}_{1'23} = \bar{\pi}_{1'}$ respectively⁴¹. The y is the same for both points. In terms of the gradient's components, $F_i = \partial F / \partial x_i$,

$$\begin{cases} x = \frac{F_2 + 2F_3}{F_1 + F_2 + F_3}, \\ y = \frac{x_1 F_1 + x_2 F_2 + x_3 F_3}{F_1 + F_2 + F_3}, \\ x' = 2 \frac{3F_1 + F_2 + 2F_3}{F_1 + F_2 + F_3}. \end{cases} \quad (9.15)$$

These transformations are the direct extension of the 2-D point \leftrightarrow point curve transformations derived in Chapter 7. A word of caution: when the interaxis distance d is not equal to 1, the right-hand sides of x and x' above need to be multiplied by d and $2d$ respectively (see (7.7)); an example is given further; see (9.22).

The generalization to the hypersurfaces \mathcal{E} of \mathbb{R}^N is direct. The image of the tangent hyperplane at a point $P \in \sigma \in \mathcal{E}$ consists of $N - 1$ points determined from (9.10) by using the appropriate axis spacing $\mathbf{d}_{\mathbf{N}}^{\mathbf{i}}$, $i = 0, \dots, N - 2$, the x coordinate being the $x_{i'}$ as given above. The resulting transformation, the N -D extension of (9.15) with N terms in the numerator and denominator, determines the point $(N - 1)$ -tuples mapping the surface $\sigma \in \mathcal{E}$ into $\bar{\sigma}$ consisting of $(N - 1)$ planar regions. Next the construction of the regions is undertaken and subsequently the *link* or *matching algorithm* between them is found.

⁴¹The short-subscripts $0'$ or $1'$ notation is used interchangeably with that showing the full index list, i.e., 123 or $1'23$.

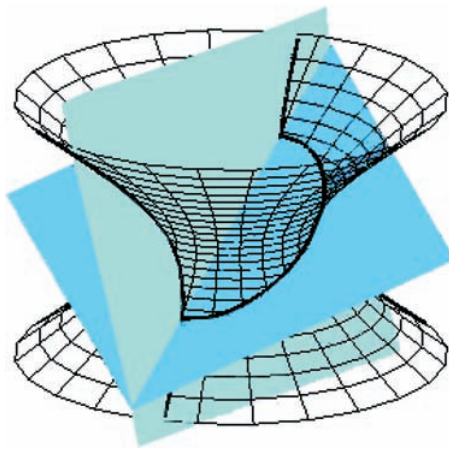


Figure 9.3. Intersection of a surface σ with the two superplanes π_0^s, π_1^s . The points of the boundary $\partial\bar{\sigma}$ are the images of the tangent planes at the points of the two curves $\sigma \cap \pi_0^s$ and $\sigma \cap \pi_1^s$. The surface above is a hyperboloid of one sheet.

9.3 **FT-2** Boundary Contours

The development is presented for \mathbb{R}^3 , where the generalization to \mathbb{R}^N , via vector notation, is straightforward. The construction of the regions $\bar{\sigma}$ and in particular their boundary $\partial\bar{\sigma}$ is greatly facilitated by an observation that is a substantial labor-saving device compared to the computation of the envelope of the tangent planes. In the ensuing, the surface σ in question is translated and rotated so that the first two hyperplanes are not tangent to the surface and their intersections are curves (i.e., not empty or isolated points), for example in Fig. 9.3. There may be many such positions and orientations, referred to as *a standard position*, and for our purposes any one will do. Once the representation of σ in a standard position is found, its representation in the original position and orientation may be found via the *translation* \leftrightarrow *rotation* duality discussed later on.

The following lemma is not true for all surfaces $\sigma \in \mathcal{E}$, as we will discover toward the end of the chapter, but it applies to a large subclass (including convex hypersurfaces), which we call \mathcal{EG} . The situation is reminiscent of the conditions leading to (6.2) for obtaining the envelope of families of curves, which is not completely general but still very useful. Roughly speaking, if $\sigma \in \mathcal{EG}$ for every tangent \bar{P} at a point $\bar{\pi}_{i'}$ of $\partial\bar{\sigma}_{i'}$, there exists a neighborhood of $\bar{\pi}_{i'}$ where all points of $\bar{\sigma}_{i'}$ are on the same side of \bar{P} . Henceforth, unless otherwise stated, when the boundaries of the representing regions of a surface σ are found using the lemma, it is assumed that $\sigma \subset \mathcal{EG}$.

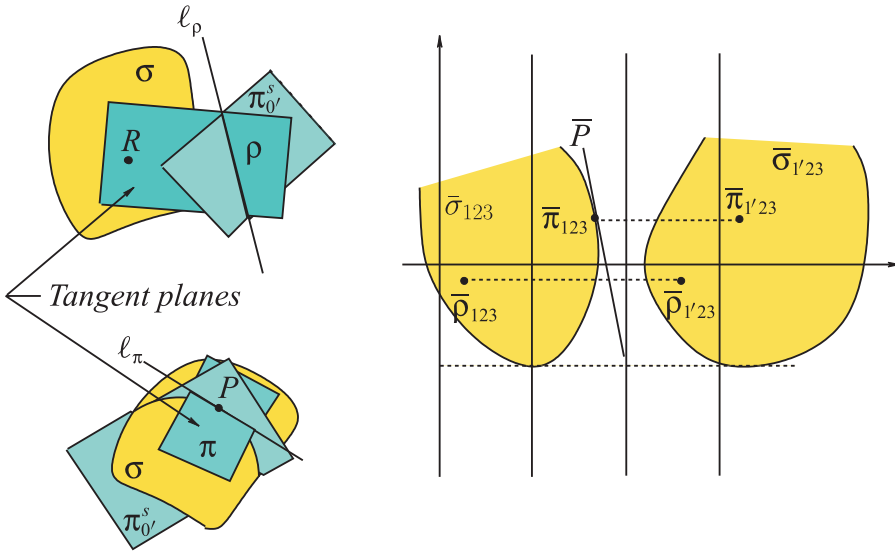


Figure 9.4. Formation of boundary contours.

Lemma 9.3.1 (Boundary of $\bar{\sigma}$) For σ a surface in \mathbb{R}^3 , $\partial\bar{\sigma}_{123}$ is the image of $\sigma \cap \pi_0^s$. That is, $\partial\bar{\sigma}_{123} = \sigma \cap \pi_0^s$ and similarly $\partial\bar{\sigma}_{1'23} = \sigma \cap \pi_1^s$.

Proof. With reference to Fig. 9.4, each point $\bar{\rho}_{123} \in \bar{\sigma}_{123}$ is a representing point of a plane ρ tangent to the surface σ at a point R by the definition of the region $\bar{\sigma}_{123}$. In fact, $\bar{\rho}_{123} = \bar{\ell}_\rho$, where $\rho \cap \pi_0^s = \ell_\rho$. Note that for a plane π tangent at point $P \in \sigma$, $\bar{\pi}_{123} \in \partial\bar{\sigma}_{123}$, $\ell_\pi = \pi \cap \pi_0^s$ and $P \in \ell_\pi$. The image \bar{P} is a line tangent to $\partial\bar{\sigma}_{123}$ at $\bar{\pi}_{123} = \bar{\ell}_{123} \Rightarrow P \in \pi_0^s$: duality *point-of-tangency* \leftrightarrow *line-of-tangency*. In short, every point of $\bar{\sigma}_{123}$ represents a line of π_0^s due to the smoothness of $\partial\sigma$ and, therefore, by continuity of $\partial\bar{\sigma}_{123}$, every tangent to $\partial\bar{\sigma}_{123}$ represents a point of π_0^s and $\partial\sigma$. Hence $\partial\bar{\sigma}_{123} = \overline{\sigma \cap \pi_0^s}$ and similarly $\partial\bar{\sigma}_{1'23} = \overline{\sigma \cap \pi_1^s}$. The smoothness of $\partial\sigma$ ensures the continuity of $\partial\bar{\sigma}_{123}$. ■

This basic result tremendously simplifies matters. The boundary $\partial\bar{\sigma}$ is simply the image of two *curves*: the intersections of σ with the first two superplanes π_0^s, π_1^s . Note further that for a plane π tangent at a point $P \in \sigma$, $\bar{\pi}_{123} \in \partial\bar{\sigma}_{123}$ and $\bar{\pi}_{1'23} \in \partial\bar{\sigma}_{1'23} \Leftrightarrow P \in \sigma \cap \pi_0^s \cap \pi_1^s$. Finding the boundary of surface representations is one of the major reasons for studying the transform of curves to $\|\text{-coords}$ in Chapter 7.

Lemma 9.3.2 (Boundary of $\bar{\sigma}$ in \mathbb{R}^N) For a $\sigma \subset \mathbb{R}^N$, $\partial\bar{\sigma}$ is composed of $N - 1$ curves that are the images of the intersections of σ with the first $N - 1$ superplanes.

Corollary 9.3.3 (Ideal points) *An ideal point on the boundary $\partial\bar{\sigma}$ is the image of a tangent plane at a point of σ , parallel to \mathbf{u} .*

Proof. It is an immediate consequence of (9.10) that ideal points occur in the image when the third component is zero. That is, the tangent plane's normal is normal to \mathbf{u} or the tangent plane is parallel to \mathbf{u} . ■

♣ **FT-2e**

An algebraic surface is one described by a polynomial equation, providing an important special case of Lemma 9.3.2.

Corollary 9.3.4 (Boundary of an algebraic surface) *The boundary $\bar{\sigma}$ of an algebraic surface σ of degree n is composed of $N - 1$ algebraic curves each of degree $n(n - 1)$ or less.*

The corresponding result in algebraic geometry⁴² is that the dual of a nonsingular algebraic surface of degree n has degree $n(n - 1)^{(N-1)}$ [52], [186]. Here the boundary representations can be found with the aid of *Plücker's* results, presented in Section 7.6. For F in (9.15) a quadratic polynomial, the corresponding surface is called a *quadric*, and in \mathbb{R}^3 ,

$$\sigma : F(x_1, x_2, x_3) = (x_1, x_2, x_3, 1)\mathbf{A}(x_1, x_2, x_3, 1)^T, \quad (9.16)$$

\mathbf{A} being a symmetric 4×4 matrix for \mathbb{R}^3 and $(N + 1) \times (N + 1)$ for \mathbb{R}^N .

Corollary 9.3.5 (Boundary of $\bar{\sigma}$ for σ a quadric) *A quadric surface $\sigma \subset \mathbb{R}^N$ is represented by $N - 1$ linked planar regions with conic boundaries.*

This is the direct extension from conics in 2-D to quadrics in N -D, and as for conics, their type depends on the type of quadric and its orientation, as pointed out next.

Incidentally, quadrics may be classified according to their matrix as in [126, pp. 455–7]. A matrix is called *positive* or *negative definit* if all its eigenvalues are positive or negative, respectively, and *indefinit* if it has both positive and negative eigenvalues. The corresponding surface has a local minimum/maximum if A is positive/negative definite and a *saddle* if A is indefinite. Numerically, a fast test for the positive definiteness of a symmetric matrix A is to attempt to factor it as $A = R^T R$, where R is upper triangular with positive diagonal entries. For a nice treatment of quadrics and their properties see also [183, Chap. VII]. Another beautiful but nonelementary classification of quadrics is in [13, vol. II, Chap. 15].

The boundary $\partial\bar{\sigma}$ is the image of *space* curves contained in the superplanes. A space curve (see also Section 7.4.1) $c \subset \mathbb{R}^N$ can also be represented as the image of its tangent lines. In turn, the tangent t at a point $P \in c$ can be represented by

⁴²I am indebted to Prof. Evgeni Shustin for these references.

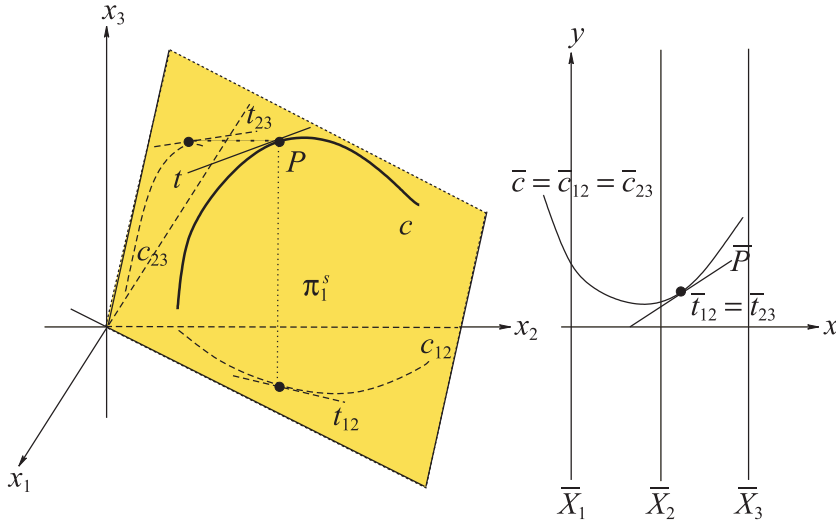


Figure 9.5. The image \bar{c} of a space curve c contained in π_1^s is formed by two coincident curves.

$N - 1$ doubly indexed points, say $\bar{t}_{i-1,i}$, such that \bar{c} consists of $N - 1$ planar curves $\bar{c}_{i-1,i}$ formed by the $\bar{t}_{i-1,i}$, two such curves for \mathbb{R}^3 . Each $\bar{c}_{i-1,i}$ is the image of the projection of c on the 2-plane $x_{i-1,i}$. The fun part is that in our case, these two curves are coincident, since they are the images of a curve in a superplane, say $c \subset \pi_{0'}^s$, where each tangent t at a point $P \in c$ is a line in the $sp \pi_{0'}^s$, and hence $\bar{t}_{12} = \bar{t}_{23}$. An example is in order. Consider the cylinder

$$CY: x_2^2 + x_3^2 = 1, \quad (9.17)$$

which is centered along the x_1 axis. For a moment we anticipate a result from the next subsection that the regions representing a cylinder are *curves* and find the representation curves starting with \overline{CY}_{12} . Based on the aforementioned, the intersection c with the first superplane

$$\pi_{0'}^s: x_1 - 2x_2 + x_3 = 0 \quad (9.18)$$

must be found. The intent is to show that the projections of $c = \pi_{0'}^s \cap CY$ on the principal planes have identical images. Substituting

$$x_3 = 2x_2 - x_1, \quad x_2 = \frac{1}{2}(x_1 + x_3),$$

successively into (9.17), we obtain

$$c_{12} = x_1^2 - 4x_1x_2 + 5x_2^2 - 1 = 0, \quad c_{13} = x_1^2 + 2x_1x_3 + 5x_2^2 - 4 = 0, \quad (9.19)$$

which are the projections of c on the x_1x_2 , x_1x_3 planes, respectively; c_{23} is available directly from (9.17). Of course, all these are planar ellipses, and we refer to Section 7.5, on the conic transforms, to find their images. For convenience the appropriate notation and formulae are restated. The image of the quadratic

$$F(x_1, x_2) = A_1x_1^2 + 2A_4x_1x_2 + A_2x_2^2 + 2A_5x_1 + 2A_6x_2 + A_3 = 0 \quad (9.20)$$

is the quadratic

$$f(x, y) = a_1x^2 + 2a_4xy + a_2y^2 + 2a_5x + 2a_6y + a_3 = 0 \quad (9.21)$$

with the coefficients

$$\left. \begin{aligned} a_1 &= [A_3(A_1 + A_2 + 2A_4) - (A_5 + A_6)^2]/d^2, \\ a_2 &= A_1A_2 - A_4^2, \\ a_3 &= A_2A_3 - A_6^2, \\ a_4 &= [A_{13} - A_{23}]/d = [A_6(A_1 + A_4) - A_5(A_2 + A_4)]/d, \\ a_5 &= [A_6^2 + A_5A_6 - A_3(A_2 + A_4)]/d, \\ a_6 &= A_2A_5 - A_4A_6, \end{aligned} \right\} \quad (9.22)$$

with d the interaxis distance and $\Delta = (A_4^2 - A_1A_2)$ the *discriminant*; the *indicator* $I = A_1 + A_2 + 2A_4$ plays an important role in the transformation. For c_{12} , the interaxis distance is $d = 1$, and the nonzero coefficients are

$$A_1 = 1, \quad A_4 = -2, \quad A_2 = 5, \quad A_3 = -1.$$

So from (9.22), the coefficients of its image are

$$a_1 = -2, \quad a_2 = 1, \quad a_3 = -5, \quad a_4 = 0, \quad a_5 = 3, \quad a_6 = 0$$

for the image curve

$$\bar{c}_{12} : -2x^2 + y^2 + 6x - 5 = 0. \quad (9.23)$$

Continuing for c_{13} , the distance between the \bar{X}_1 and \bar{X}_3 axes is $d = 2$, the nonzero coefficients (where now the second variable is x_3) and its image's coefficients are

$$\begin{aligned} A_1 &= 1, \quad A_4 = 1, \quad A_2 = 5, \quad A_3 = -4, \quad a_1 = -8, \quad a_2 = 4, \\ a_3 &= -20, \quad a_4 = 0, \quad a_5 = 24, \quad a_6 = 0, \end{aligned}$$

which, upon division by 4, define the same curve as in (9.23). Note that without the insertion of the interaxis distance d in the formulas, the computation would

not be correct. The image \bar{c}_{23} is found next to verify the assertion. The nonzero coefficients of c_{23} and its image are

$$\begin{aligned} A_1 &= 1, & A_2 &= 1, & A_3 &= -1, & a_1 &= -8, & a_2 &= 1, & a_3 &= -1, \\ a_4 &= 0, & a_5 &= 1, & a_6 &= 0, \end{aligned}$$

with the distance between the \bar{X}_2 and \bar{X}_3 axes $d = 1$. We write

$$\bar{c}_{23} : -2x'^2 + y^2 + 2x' - 1 = 0, \quad x' = x - 1, \quad (9.24)$$

since the \bar{X}_2 axis is one unit to the right of the y axis. In the xy coordinate system,

$$\bar{c}_{23} : -2x^2 + y^2 + 6x - 5 = 0. \quad (9.25)$$

Hence images of all the projections of c , which are ellipses, are coincident hyperbolas $\bar{c}_{12} = \bar{c}_{13} = \bar{c}_{23} = \bar{c} = \overline{CY}_{12}$.

Proceeding, the image \overline{CY}_{23} is the image of the curve $c' = CY \cap \pi_{1'}^s$, where

$$\pi_{1'}^s : x_1 + x_2 - 2x_3 = 0 \quad (9.26)$$

is the second sp based on the coordinate axis $\bar{X}_{1'}, \bar{X}_2, \bar{X}_3$. Actually $\overline{CY}_{23} = \bar{c}_{23}$ has already been obtained above, since the \bar{X}_2, \bar{X}_3 axes are shared by the two coordinate systems. This means that the two parts $\overline{CY}_{23}, \overline{CY}_{12}$ of \overline{CY} are coincident, an occurrence that turns out to be characteristic of cylinders whose rulings are parallel to the x_1 axis. It is instructive to verify this in terms of the other two projections of c' found by the substitutions from (9.26),

$$x_2 = 2x_3 - x_1, \quad x_3 = \frac{1}{2}(x_1 + x_2),$$

into (9.17), which yields

$$c'_{1'3} : x_1^2 - 4x_1x_3 + 5x_3^2 - 1 = 0, \quad c'_{1'2} : x_1^2 + 2x_1x_2 + 5x_2^2 - 4 = 0. \quad (9.27)$$

The nonzero coefficients of $c'_{1'3}$ and the a_i are

$$\begin{aligned} A_1 &= 1, & A_4 &= -2, & A_2 &= 5, & A_3 &= -1, & a_1 &= -2, & a_2 &= 1, \\ a_3 &= -5, & a_4 &= 0, & a_5 &= -3, & a_6 &= 0, \end{aligned}$$

with $d = -1$ the (directed) distance between the $\bar{X}_{1'}, \bar{X}_3$ axes, yielding

$$\bar{c}'_{1'3} : -2x'^2 + y^2 - 6x' - 5 = 0.$$

Here $x' = x - 3$ due to the distance of three units of $\bar{X}_{1'}$ from the y axis, so

$$\bar{c}'_{1'3} : -2(x - 3)^2 + y^2 - 6(x - 3) - 5 = -2x^2 + y^2 + 6x - 5 = 0, \quad (9.28)$$

which matches \bar{c}_{23} . Finally, $d = -2$ and $x' = x - 1$ for $c'_{1'2}$, whose image is also the above hyperbola.

This extended presentation illustrates the detail involved and emphasizes the simplification achieved. For each region, the image of *only one* projection needs to be computed, so we may as well pick the simplest one, where the two parallel axes are adjacent with $d = 1$. The conclusion, a refinement of Lemma 9.3.2, is stated in general.

Theorem 9.3.6 (Representation of $\bar{\sigma}$ in \mathbb{R}^N) *A smooth hypersurface $\sigma \subset \mathbb{R}^N$ can be represented by $N - 1$ linked (i.e., with a matching algorithm) regions $\bar{\sigma}_{i,i+1} \subset \mathbb{P}^2, i = 1, \dots, N - 1$, with $\partial \bar{\sigma}_{i,i+1} = \bar{c}_{(i'-1),i'} = (\sigma \cap \pi_{i'}^{Ns})_{(i'-1),i'}$.*

To clarify the notation, $\pi_{i'}^{Ns}$ is the $(i + 1)$ th *sp* (recall that the first one is also denoted by π_0^{Ns}). The image of the *space curve* $\pi_{i'}^{Ns} \cap \sigma$ is the image of *only one* projection; we choose $(\pi_{i'}^{Ns} \cap \sigma)_{(i'-1),i'}$ corresponding to the adjacent axes $\bar{X}_{i'-1}, \bar{X}_{i'}$ for which $d = 1$, with the convention that for $i' = 1, \bar{X}_{(i'-1)} = \bar{X}_N$. A horizontal translation by $N + i - 1$ units to the left as shown in Fig. 9.6 is needed to obtain the correct x coordinate. For an object, a point P , for example, the notation $\bar{P}_{i'}$ refers to the representation of P with respect to the N coordinate axes after the i th shift.

For \mathbb{R}^3 the preferred projections are $(\pi_1^s \cap \sigma)_{12}, (\pi_1^s \cap \sigma)_{1'3}$, and for \mathbb{R}^4 , $(\pi_1^{4s} \cap \sigma)_{12}, (\pi_1^{4s} \cap \sigma)_{1'4}, (\pi_1^{4s} \cap \sigma)_{1'2'}$, and so on. Next, the regions of $\bar{\sigma}$ must be *linked* with a *matching algorithm*, as for the regions in Chapter 8 representing families of proximate flats, which select the valid $N - 1$ points representing a tangent plane of σ . This is dealt with incrementally for surface types of increasing complexity.

9.4 FT-3 Developable Surfaces

The sphere cannot be cut and then flattened without distortion. Motivated by mapmaking, surfaces were sought whose shape is “close” to spherical and can be unrolled into a plane without stretching or contracting. Euler first considered this problem, and subsequently, starting in 1771, Monge made major contributions to the subject of *developable* surfaces. Monge pointed out potential applications especially to architecture, paving the way for the modern contoured architectural marvels. Gauss and others followed with the development of the differential geometry of

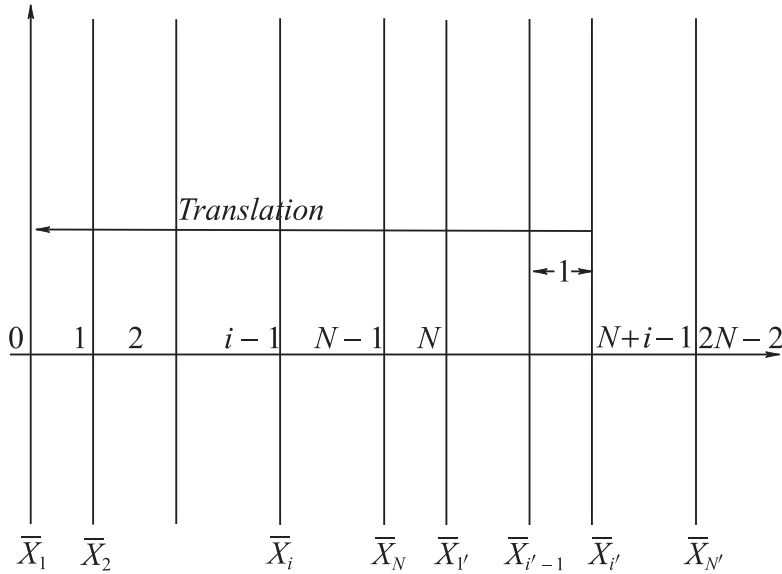


Figure 9.6. Notation and distances for the shift by i units in \mathbb{R}^N .

Due to the distance of the $\bar{X}_{i'}$ from the y axis, the horizontal translation by $N - i + 1$ is needed to obtain the correct x coordinate.

more genenal surfaces. Developable surfaces (“developables” for short) are the class $\mathcal{D} \subset \mathcal{E}$ that are the envelope of a one-parameter family of planes; they serve as an excellent starting point for our study of surface representation. Finding their image, matching, and reconstruction algorithms is straightforward, and the results offer crucial guides for coping with the more general representation problems. The pioneering work, and the basis for the exposition here on their representation, is due to C.K. Hung [93], [92], and the recent more complete exposition [94]. Formally, let the family of tangent planes describing a developable σ be specified by

$$\pi(t) : \mathbf{c}(t) \cdot \mathbf{x} = c_0(t), \quad t \in I_t \subset \mathbb{R}, \quad (9.29)$$

where I_t is an interval of \mathbb{R} . The conditions of (9.35) are added, as proved below, to guarantee that the corresponding developable is well defined. In other words, it is not a single plane, a family of parallel planes, nor a pencil of planes intersecting on a line.

Consider the two planes $\pi(t) : \mathbf{c}(t) \cdot \mathbf{x} = c_0(t)$, $\pi(t + \Delta t) : \mathbf{c}(t + \Delta t) \cdot \mathbf{x} = c_0(t + \Delta t)$ intersecting at a line r , which is necessarily also on the plane

$$[\mathbf{c}(t + \Delta t) - \mathbf{c}(t)] \cdot \mathbf{x} = [c_0(t + \Delta t) - c_0(t)].$$

Dividing by Δt and taking the limit as $\Delta t \rightarrow 0$, the limiting position of r is

$$r(t) : \mathbf{c}(t) \cdot \mathbf{x} = c_0(t), \quad \mathbf{c}'(t) \cdot \mathbf{x} = c'_0(t), \quad (9.30)$$

the prime $'$ denoting differentiation with respect to the parameter t . This line is called a *ruling*, *generator*, or *characteristic* of the plane $\pi(t)$ that contains it. Three planes $\pi(t)$, $\pi(t + \Delta_1 t)$, $\pi(t + \Delta_1 t + \Delta_2 t)$ intersect at a point C , which is also on the plane

$$[\mathbf{c}(t + \Delta_1 t + \Delta_2 t) - \mathbf{c}(t + \Delta_2 t)] \cdot \mathbf{x} = [c_0(t + \Delta_1 t + \Delta_2 t) - c_0(t + \Delta_1 t)].$$

As above, dividing by $\Delta_2 t$ and taking the limit as $\Delta_2 t \rightarrow 0$, we obtain the first of the planes

$$\pi_d(t + \Delta_1 t) : \mathbf{c}'(t + \Delta_1 t) \cdot \mathbf{x} = c'_0(t + \Delta_1 t), \quad \pi_d(t) : \mathbf{c}'(t) \cdot \mathbf{x} = c'_0(t),$$

with the second one obtained similarly. Forming their difference, dividing by $\Delta_1 t$, and taking the limit as $\Delta_1 t \rightarrow 0$, we find the limiting position of the point C as the intersection of the three planes

$$\mathbf{c}(t) \cdot \mathbf{x} = c_0(t), \quad \mathbf{c}'(t) \cdot \mathbf{x} = c'_0(t), \quad \mathbf{c}''(t) \cdot \mathbf{x} = c''_0(t). \quad (9.31)$$

This is the *characteristic point* of the plane $\pi(t)$; it lies on the plane's ruling $r(t)$. When all the $\pi(t)$ enveloping a developable σ have the same characteristic point, σ is a *cylinder*, whose rulings are parallel if the point is ideal, or a *cone* when all rulings lie on a regular point.

We state without proof that in general, the locus of characteristic points is a curve cc , called the *edge of regression*, and at each point $C(t) \in cc$ the ruling $r(t)$ of the same plane $\pi(t)$ is tangent to cc . The family of planes (9.29) satisfying the conditions in (9.35) has an envelope that is a cylinder, a cone, or a *tangential developable* having two sheets generated by the rulings tangent to the curve cc . It is the envelope of its osculating planes. The intersection of the tangential developable with the plane normal to the ruling $r(t)$ at a point $C(t)$ has a cusp at that point. Induced by the *point* \leftrightarrow *plane* duality, there is an interesting duality between space curves, considered as the locus of a one-parameter family of points and developable surfaces formed by a one-parameter family of tangent planes [166, p. 72]. This can serve as a blueprint for an analogous duality in \mathbb{R}^N , which we do not pursue here. Without entering into further detail, we mention that a necessary and sufficient condition for a smooth surface $\sigma \in \mathcal{E}$ to be developable is that its *Gaussian curvature* vanishes [166, pp. 73–91]. A specific example, the circular double-cone, helps clarify matters. The parametric equations describing the surface are derived with the aid of Fig. 9.7, showing a cone CC symmetrical about the x_3 axis with apex angle α . Let $P \in CC$

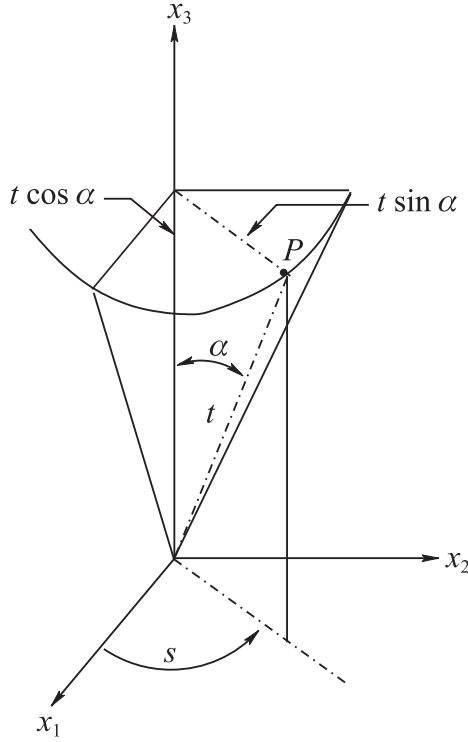


Figure 9.7. Deriving the parametric equations of a circular cone.

be a point at a distance t from the apex on the circular slice of CC with the plane $x_3 = t \cos \alpha$. The radius of P from the x_3 axis is $t \sin \alpha$ and at an angle s from the x_1 axis. The coordinates of P in terms of the two parameters (s, t) are

$$CC : \mathbf{x}(s, t) = (x_1 = t \sin \alpha \cos s, x_2 = t \sin \alpha \sin s, x_3 = t \cos \alpha). \quad (9.32)$$

Next we derive the family of tangent planes of CC using (9.12):

$$\pi : \begin{vmatrix} x_1 - x_1^0 & x_2 - x_2^0 & x_3 - x_3^0 \\ -t \sin \alpha \sin s & t \sin \alpha \cos s & 0 \\ \sin \alpha \cos s & \sin \alpha \sin s & \cos \alpha \end{vmatrix} = 0. \quad (9.33)$$

The parameter t in the second row cancels out, and so the developable CC has a *one-parameter* family of tangent planes

$$\pi(s) : (x_1 - x_1^0) \cos \alpha \cos s + (x_2 - x_2^0) \cos \alpha \sin s - (x_3 - x_3^0) \sin \alpha = 0. \quad (9.34)$$

A one-parameter family of planes is mapped into two sets of points, a pair for each parameter value, so the two regions representing developables are *curves*. Though this is intuitively clear, the rigorous proof below, needed to exclude pathological cases, is involved, and its reading can be omitted.

Theorem 9.4.1 (C.K. Hung; Developable Surfaces) *Let a surface $\sigma \in \mathcal{D}$ with tangent planes given by (9.29), and a neighborhood $U \subset I_t \subset \mathbb{R}$, where the two conditions*

$$\frac{dx(t)}{dt} = \frac{\partial}{\partial t} \left(\frac{\mathbf{c}(t) \cdot \mathbf{d}^i}{\mathbf{c}(t) \cdot \mathbf{u}} \right) = 0, \quad \frac{dy(t)}{dt} = \frac{\partial}{\partial t} \left(\frac{c_0(t)}{\mathbf{c}(t) \cdot \mathbf{u}} \right) = 0, \quad (9.35)$$

are not simultaneously satisfied for all $t \in U$. Then the set of points $\{\bar{\pi}(t)_{i'} | t \in U\}$ representing the tangent planes (9.29) are curves.

♣ FT-3e

**** Proof.** The points representing the planes (9.29) specifying the developable surface σ are

$$\bar{\pi}(t)_{i'} = (\mathbf{c}(t) \cdot \mathbf{d}^i, \quad c_0(t), \quad \mathbf{c}(t) \cdot \mathbf{u}) \quad \forall t \in I_t. \quad (9.36)$$

These points are the image of a curve defined by the parameter $t \in I_t$, provided that the functions $x(t)$, $y(t)$ are single-valued and the conditions (9.35) are satisfied. These conditions fail if and only if there exist constants k_1, k_2 such that

$$\begin{cases} \mathbf{c}(t) \cdot \mathbf{d}^i \equiv k_1 \mathbf{c}(t) \cdot \mathbf{u}, \\ c_0(t) \equiv k_2 \mathbf{c}(t) \cdot \mathbf{u}, \end{cases} \Rightarrow \begin{cases} \mathbf{c}(t) \cdot (\mathbf{d}^i - k_1 \mathbf{u}) \equiv 0, \\ \mathbf{c}(t) \cdot (k_2 \mathbf{u}) \equiv c_0(t), \end{cases} \quad (9.37)$$

identically for t in the neighborhood $U \subset I_t$. Consider first the case $k_2 = 0$. Then necessarily $c_0(t) \equiv 0 \Rightarrow$ that all the planes $\pi(t)$ contain the origin. Let $\hat{\mathbf{e}}_1$ and ℓ_1 denote the unit vector and line respectively through the origin in the direction of $(\mathbf{d}^i - k_1 \mathbf{u})$, i.e.,

$$\hat{\mathbf{e}}_1 = \frac{(\mathbf{d}^i - k_1 \mathbf{u})}{|\mathbf{d}^i - k_1 \mathbf{u}|} = K_1 (\mathbf{d}^i - k_1 \mathbf{u}).$$

Then the first identity implies that $\forall t \in U$, $\hat{\mathbf{e}}_1$ is orthogonal to the vector $\mathbf{c}(t)$ which is normal to the plane $\pi(t)$. Unless $\mathbf{c}'(t) \equiv 0$ for all $t \in U$, then

$$\bigcap_{t \in U} \pi(t) = \ell_1, \quad (9.38)$$

showing that the family of planes $\pi(t)$ given by (9.29), rather than being the envelope of a developable surface, is either a single plane (and hence σ itself) or a pencil of planes on the line ℓ_1 .

Next consider the case $k_2 \neq 0$. From the first identity, $\hat{\mathbf{e}}_1 \perp \mathbf{c}(t)$, and from the second, $\mathbf{c}(t) \not\perp k_2 \mathbf{u} \Rightarrow \hat{\mathbf{e}}_1$ and $k_2 \mathbf{u}$ must be linearly independent; for if not, then they would be in the same direction and hence both would be orthogonal to $\mathbf{c}(t)$. So every linear combination with nonzero coefficients of the two vectors is not zero, and in particular, $\mathbf{u} - (\mathbf{u} \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1 \neq 0$. Let us denote the unit vector in this direction by $\hat{\mathbf{e}}_2$, i.e.,

$$\hat{\mathbf{e}}_2 = \frac{(\mathbf{u} - (\mathbf{u} \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1)}{|\mathbf{u} - (\mathbf{u} \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1|} = K_2[\mathbf{u} - (\mathbf{u} \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1].$$

The two unit vectors are orthogonal, since $\hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_2 = K_2 \hat{\mathbf{e}}_1 \cdot (\mathbf{u} - (\mathbf{u} \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1) = 0$. Further, $\mathbf{c}(t) \cdot \hat{\mathbf{e}}_2 = K_2 \mathbf{c}(t) \cdot \mathbf{u}$. Then (9.37) reduces to

$$\begin{cases} \mathbf{c}(t) \cdot \hat{\mathbf{e}}_1 \equiv 0, \\ \mathbf{c}(t) \cdot \hat{\mathbf{e}}_2 \equiv K c_0(t), \end{cases} \quad (9.39)$$

with K constant. Recall the vector identities

$$\begin{cases} (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = (\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}), \\ (\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{c} \cdot \mathbf{a})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{a}. \end{cases} \quad (9.40)$$

We verify by direct substitution that the vector

$$\mathbf{c}(t) = K c_0(t) \hat{\mathbf{e}}_2 + g(t)(\hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2) \quad (9.41)$$

satisfies (9.39) for an arbitrary function $g(t) \in C^1$. To wit,

$$\begin{cases} \mathbf{c}(t) \cdot \hat{\mathbf{e}}_1 = K c_0(t) \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_1 + g(t)(\hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2) \cdot \hat{\mathbf{e}}_1 = 0, \\ \mathbf{c}(t) \cdot \hat{\mathbf{e}}_2 = K c_0(t) \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_2 + g(t)(\hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2) \times \hat{\mathbf{e}}_2 = K c_0(t), \end{cases} \quad (9.42)$$

from the orthogonality of $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ and the vector identities. It consists of a vector in the direction of $\hat{\mathbf{e}}_2$ and another perpendicular to the direction of $\hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2$. The family of tangent planes, (9.29), can be rewritten as

$$[K c_0(t) \hat{\mathbf{e}}_2 + g(t)(\hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2)] \cdot \mathbf{x} = c_0(t). \quad (9.43)$$

From

$$\begin{cases} (K c_0(t) \hat{\mathbf{e}}_2 + g(t)(\hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2)) \cdot \hat{\mathbf{e}}_1 \equiv 0, \\ (K c_0(t) \hat{\mathbf{e}}_2 + g(t)(\hat{\mathbf{e}}_1 \times \hat{\mathbf{e}}_2)) \cdot (\hat{\mathbf{e}}_2 + s \hat{\mathbf{e}}_1) = K c_0(t), \end{cases}$$

it is clear that the line $\ell_2 : \hat{\mathbf{e}}_2 + s \hat{\mathbf{e}}_1$, parametrized by s , is contained in all the tangent planes (9.43) of σ . Again, provided that $\mathbf{c}'(t) \neq 0$, these planes, rather than

being the envelope of a surface, are a pencil of planes on the line ℓ_2 . In short, the requirement that the tangent planes of a surface $\sigma \in \mathcal{D}$ not satisfy the conditions in (9.35) ensures that (9.36) defines two (piecewise) \mathcal{C}^1 curves in \mathbb{R}^3 that represent a developable not simultaneously satisfying the two conditions of (9.35). Use of vector notation throughout allows us to conclude that the $N - 1$ curves

$$\bar{\sigma}_{i'}(t) = \{\bar{\pi}_{i'}(t) = (\mathbf{c}(t) \cdot \mathbf{d}^i, c_0(t), \mathbf{c}(t) \cdot \mathbf{u})\}, \quad i = 0, \dots, N - 2, \quad (9.44)$$

represent such a developable in \mathbb{R}^N . Since each point set $\bar{\pi}_{i'}(t)$ depends on only one parameter, it is clear that for each i , it is a curve $\bar{\sigma}_{i'}(t)$ rather than a region (i.e., it has no interior with points). That is, for σ a developable,

$$\bar{\sigma} = \{\partial \bar{\sigma}_{i'} | i = 0, \dots, N - 2\}.$$

■

9.4.1 **FT-4** Ambiguities and Uniqueness

Can a pair of curves represent different developable surfaces? Moreover, can a developable surface be *reconstructed* from its representation? These questions are interrelated, and a proper answer requires an algorithm for identifying every tangent plane of σ from the two points, one on each of the representing curves. The solution is immediate when the curves are available in parametric form $\bar{\sigma}_{123}(t)$ and $\bar{\sigma}_{231'}(t)$; each value of t provides two unique points identifying each tangent plane $\pi(t)$. The challenge arises when the representing curves $\bar{\sigma}_{123}, \bar{\sigma}_{231'}$ are given without parametrization. Recall that a plane π is uniquely identified by two points having the same y -coordinate. Therefore, if one of the representing curves is monotonic in the y -direction, so is the other. The intersection of the horizontal line through a point $\bar{\pi}_{123} \in \bar{\sigma}_{123}$ with the curve $\bar{\sigma}_{231'}$ is on the second point $\bar{\pi}_{231'}$ corresponding to the tangent plane π . In effect, then the curves can be parametrized by either x or y . Otherwise, for each monotone portion of the $\bar{\sigma}_{123}$ curve there must be a corresponding monotone portion of $\bar{\sigma}_{1'23}$ in the same range of y . If not, the curves cannot represent a developable surface. A unique correspondence can be established by a pair of *start* points $\bar{\pi}_{123} \in \bar{\sigma}_{123}, \bar{\pi}_{231'} \in \bar{\sigma}_{231'}$, on a horizontal line \bar{H} identifying a specific tangent plane π . Tracking the continuous translation of the line \bar{H} in the direction of increasing or decreasing y as in Fig. 9.8 provides the correspondence of point pairs, one on each curve, identifying the tangent planes. Between the curves' extrema, \bar{H} may intersect each curve at several points. For example, there are two points of intersection for each ellipse in Fig. 9.8, allowing for four possible pairings of start points, providing, in general, four *different* developables; and there are other problems. In the examples below, though the curves may be traversable with

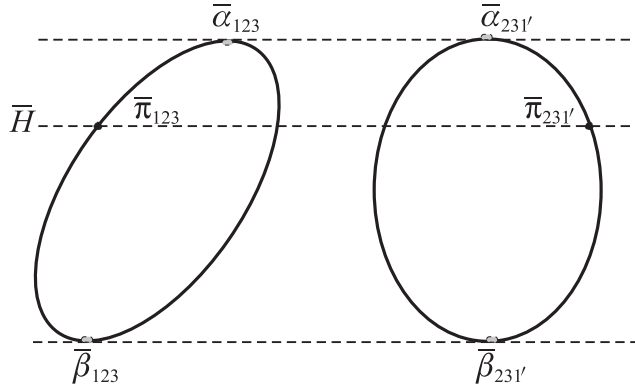


Figure 9.8. The pair of points $\bar{\pi}_{123}, \bar{\pi}_{231'}$ represents an unambiguous plane. By contrast the $\bar{\alpha}$ and $\bar{\beta}$ points represent planes which are *a1*, ambiguous in the first sense.

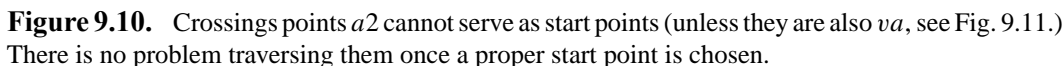
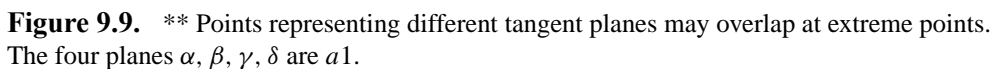
a proper partition and matching, they may not always represent a valid developable. This issue is examined later on.

♣ FT-4e

Definition (Ambiguous plane) A tangent plane π of a surface $\sigma \in \mathcal{D}$ is ambiguous in the first sense, denoted by *a1*, if either $\bar{\pi}_{123}$ or $\bar{\pi}_{1'23}$ is an extreme point in the y -direction of its corresponding curve. It is ambiguous in the second sense, denoted by *a2*, if either of the points is a crossing point on its curve. All other planes are unambiguous.

The notation *a1* or *a2* is used interchangeably either for ambiguous planes or the points representing them. Returning to Fig. 9.8, it is seen that the planes α and β are *a1*, since they are represented by extreme points of the curves. They cannot serve as start points, for in the absence of other information, the curve's traversal can proceed either to the left or right of the points. Instances of different representing points on one of the curves are shown in Fig. 9.9. Such points are necessarily at an extremum. Figure 9.10 illustrates the ambiguity, due to the multiple choice of directions, in attempting to traverse the curve starting from a crossing point. Otherwise, choosing a start point, say $\bar{\gamma}_{123}$, limits the choice of $\bar{\gamma}_{231'}$ to be on a matching monotone portion inducing the traversal in the indicated direction. Reaching the crossing point⁴³ confronts us with a choice of directions. Here the stipulation that the curves be C^1 eliminates the ambiguity, forcing the traversal to occur along the path preserving the continuity of the direction. That is, the path of traversal has no “kinks.” Another

⁴³In Fig. 9.24 it is seen that crossing points represent a bitangent plane. I am grateful to Y. Perlmutter for proving that in the representation of developables, crossing points occur in pairs; Exercise 3.



Definition (Very Ambiguous plane) *Let π be an a_2 plane such that at least one of its representing points, say $\bar{\pi}_{123}$, is a crossing point. It is called very ambiguous, and denoted by va , if in addition either*

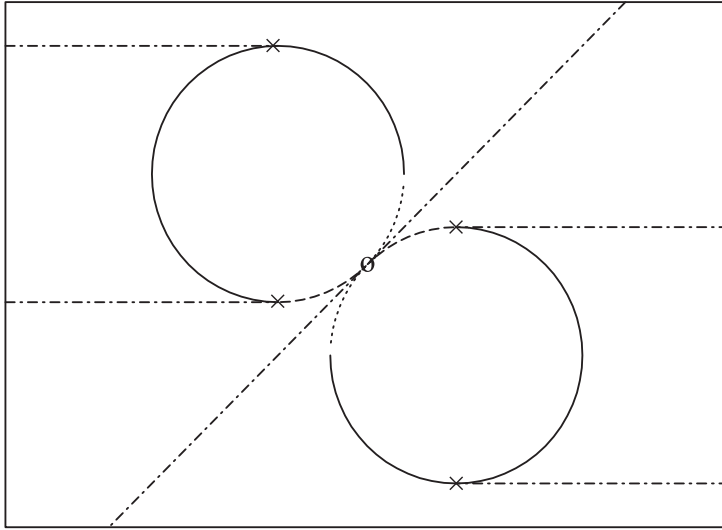


Figure 9.11. One of the curves representing a developable surface σ has four $a1$ points. They are marked by “x.” There is also an $a2$ point marked by “o,” which, being a crossing point where the tangents of the curve’s two parts coincide, is also a va .

1. $\bar{\pi}_{123}$ (and hence $\bar{\pi}_{231'}$) is not an ideal point and at least one of the curves has a pair of branches sharing the same tangent direction at $\bar{\pi}_{123}$ (or $\bar{\pi}_{231'}$), or
2. $\bar{\pi}_{123}$ (and hence $\bar{\pi}_{231'}$) is an ideal point shared by a pair of curves’ branches.

When a va point (see Figs. 9.11 and 9.12) is reached, the traversal cannot proceed, since either choice of branch is possible. Of course, a way to cope with this ambiguity is to partition the curve(s) at va points into separate branches without va points and then match the pairs as outlined earlier. The strategy is to proceed with the curves’ traversal, avoiding previously traced portions of the curve. The discussion is basically the proof of the result formally stated below.

Lemma 9.4.2 (Unique Representation of Developables) *Let σ be a smooth developable with no va tangents represented by two smooth curves $\bar{\sigma}_{123}$, $\bar{\sigma}_{1'23}$ and π (referred to as the start plane) any unambiguous tangent plane of π . Then the representing curves $\bar{\sigma}_{123}$, $\bar{\sigma}_{231'}$ together with the two points $\bar{\pi}_{123} \in \bar{\sigma}_{123}$, $\bar{\pi}_{231'} \in \bar{\sigma}_{231'}$ uniquely represent σ .*

The suggested partitioning above at va points, let us call it *patching*, can be applied at points with more than two self-intersections, again resolving ambiguities. Allowing *patching*, there always exists a unique developable surface represented by each pair of the resulting *patched* curves and good start points.

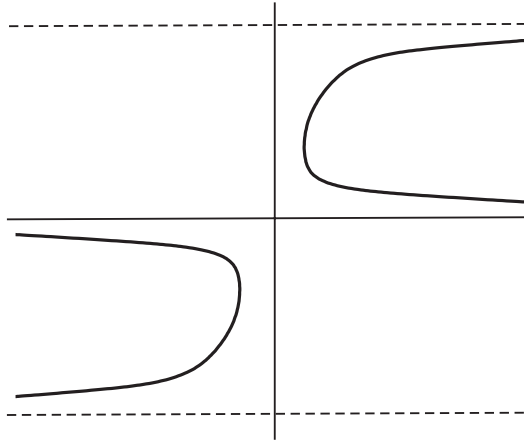


Figure 9.12. ** Two branches whose asymptotes are all horizontal. The ideal point in the horizontal direction is a *va*.

9.4.2 Reconstruction of Developable Surfaces

Henceforth, when we refer to the representation $\bar{\sigma}$ of a developable, the curve portions under discussion are assumed smooth and monotone. That is, $\bar{\sigma}$ is partitioned and matched into portions that can be uniquely traversed such as $\bar{\sigma}_{123}$ and $\bar{\sigma}_{1'23}$ seen in Fig. 9.13, with which we illustrate the reconstruction of the developable σ ; reference also to Fig. 9.4 is helpful. The points $\bar{\pi}_{123}$, $\bar{\pi}_{1'23}$ with the same y represent a tangent plane of the developable σ . The tangent lines \bar{P} , \bar{Q} represent the points

$$P \in \pi \cap \pi_{0'}^s \cap \sigma, \quad Q \in \pi \cap \pi_{1'}^s \cap \sigma,$$

so that the line $r = \pi \cap \sigma$ on P , Q is a ruling of the developable with

$$\bar{r}_{23} = \bar{P} \cap \bar{Q}, \quad \bar{r}_{12} = \bar{Q}_0 \cap \bar{P}, \quad (9.45)$$

where \bar{Q}_0 denotes the representation of Q in the first coordinate system $\bar{X}_1, \bar{X}_2, \bar{X}_3$ (recall the notation after Theorem 9.3.6).

The representation of r is based on the construction discussed in Section 5.3.1 for determining that a line is contained in a plane. To repeat, the point P on the ruling r is on σ and $\pi_{0'}^s$ and $Q \in r \cap \pi_{1'}^s$. All points on this ruling (and σ) are constructed in the usual way for lines, i.e., choosing a value of one variable and constructing the remaining coordinates from the points $\bar{r}_{12}, \bar{r}_{23}$.

Theorem 9.4.3. *A developable with a well-matched representation is (piecewise) reconstructible.*

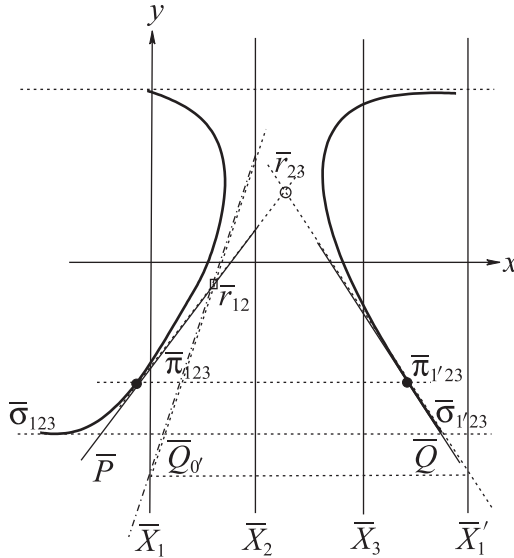


Figure 9.13. Reconstructing a developable.

The ruling r is represented by $\bar{r}_{23} = \bar{P} \cap \bar{Q}$, shown by a \circ , and $\bar{r}_{12} = \bar{P} \cap \bar{Q}_0$, shown by a \square , where \bar{Q}_0 is the 1, 2 part of the representation of Q with respect to the coordinate axes $\bar{X}_1, \bar{X}_2, \bar{X}_3$.

Other rulings are constructed in the same way: choosing a value y that determines the pair of points representing a tangent plane, finding the tangents, etc. Conversely, given the two representing curves of a developable σ , a line ℓ is one of its rulings if and only if there are two lines on $\bar{\ell}_{23}$ tangent to $\bar{\sigma}_{123}$ and $\bar{\sigma}_{1'23}$ at points with the same y . An example of a line that is not a ruling is seen in Fig. 9.14.

What is the locus $\bar{r}\bar{r}_{23}$ of all points \bar{r}_{23} where r is a ruling of a developable? This is obtained as the result of performing the above construction for all y in the representation of σ . The intersection of the two tangents at the pair of points with the same y representing a tangent plane for all y needs to be found (Exercise 7). Conversely, given the two loci, construct the developable from its rulings (see Exercise 8). There follows a pleasing result on the rulings of a developable with reflected representing curves.

Corollary 9.4.4 (Atariah, Biniaminy, Itai, Kosloff (ABIK); Rulings of reflected representing curves) *If a developable σ is represented by two curves $\bar{\sigma}_0, \bar{\sigma}_{1'}$, one the reflectio of the other about a line $x = a$, then the rulings' representing curves are two vertical lines with $\bar{r}\bar{r}_{23}$ at $x = a$ and $\bar{r}\bar{r}_{12}$ at $x = 3 - 2a$.*

The proof is left as an exercise (see Exercise 6 below).

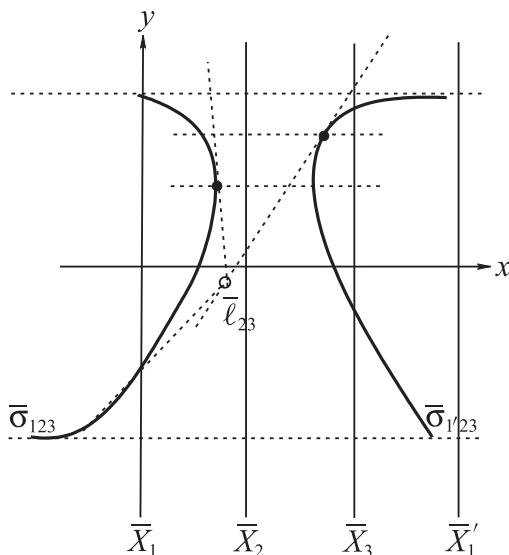


Figure 9.14. The line ℓ with the $\bar{\ell}_{23}$ shown cannot be a ruling of the developable σ . This is because the lines on $\bar{\ell}_{23}$ have points of tangency to $\bar{\sigma}_{123}$ and $\bar{\sigma}_{1'23}$ with different y .

Let us look at a couple of examples⁴⁴ that also illustrate the reflected representing curves property. Figure 9.16 (left) shows two cubic curves, one the reflection of the other about a vertical line, representing a developable σ whose rulings are represented by two vertical lines (right). The developable σ is reconstructed from its rulings and seen in Fig. 9.17. The line of cusps is the dual of the inflection points in the representing curves. It is seen and discussed further later on. Another example is shown in Fig. 9.18, where the representing curves (right) are two exponentials where there is no symmetry with respect to a vertical line. The curves representing the rulings are shown on the (right) and the developable in Fig. 9.19.

Points on the rulings are necessarily surface points of the developable. This raises the question of how to recognize on which “side” of the developable an arbitrary point lies, in particular, when the developable is closed, how to construct interior points. This is answered partially, starting with cylinders, in the next section.

⁴⁴I am grateful to my students D. Atariah, M. Biniaminy, T. Itai, and Z. Kosloff for providing these examples and results, including the corollary above from their course project.

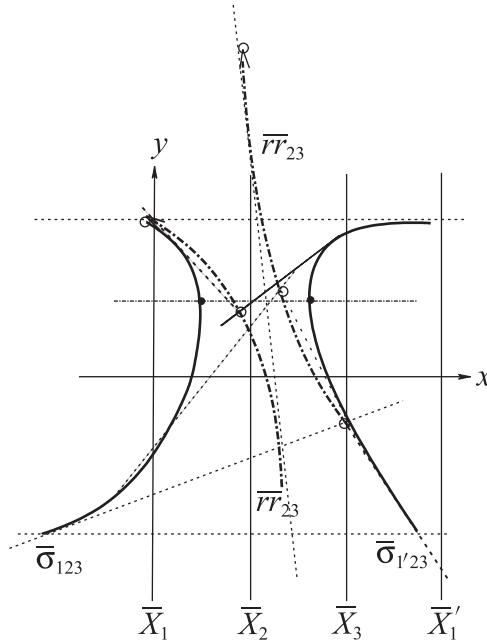


Figure 9.15. For the rulings r the locus of points \bar{r}_{23} , constructed as in Fig. 9.13.

This is a curve \bar{r}_{23} . At the pair of points marked by \bullet the tangents are vertical, resulting in an ideal point splitting \bar{r}_{23} . The arrow's direction indicates the y coordinate of \bar{r}_{23} increases monotonically with the y of the corresponding points representing to the tangent plane.

Exercises

- Which of the following ambiguities are “resolvable” and how?
 - representation curves have one horizontal segment.
 - representation curves have multiple crossings at the same point.
- For an ambiguous plane π can *only* one of $\bar{\pi}_{123}$, $\bar{\pi}'_{231}$ be an extremum but not the other?
- Give an example of a pair of curves that are not C^1 and cannot uniquely represent a developable, and another such pair that uniquely represent a developable surface.
- In the representation $\bar{\sigma}_{123}$, $\bar{\sigma}'_{1'23}$ of a developable σ prove that for a ruling r , $\bar{r}_{12} = \bar{r}_{23}$, where at the corresponding pair $\bar{\pi}_{123}$, $\bar{\pi}'_{1'23}$ one of the tangents is horizontal. Hint: see the highest pair of points on the two curves in Fig. 9.15.
- Prove that the point of intersection $\bar{\sigma}_0 \cap \bar{\sigma}_{1'}$ is the \bar{r}_{23} of a ruling r of the developable σ , e.g., the intersection of the two curves in Fig. 9.18 is such a point.

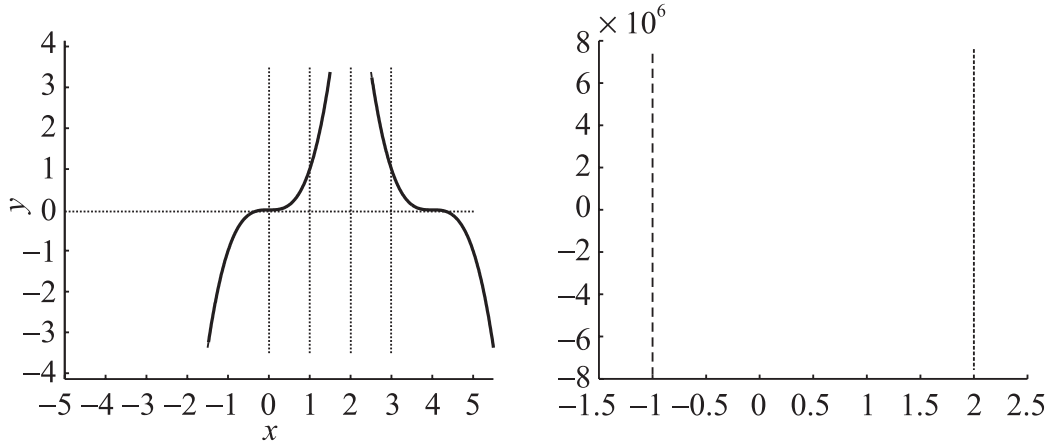


Figure 9.16. Two cubic curves symmetric about a line $x = a$ representing a developable. On the right are the loci representing the developable's rulings; two vertical lines with $\overline{r}r_{23}$ at $x = a$ and $\overline{r}r_{12}$ at $x = 3 - 2a$.

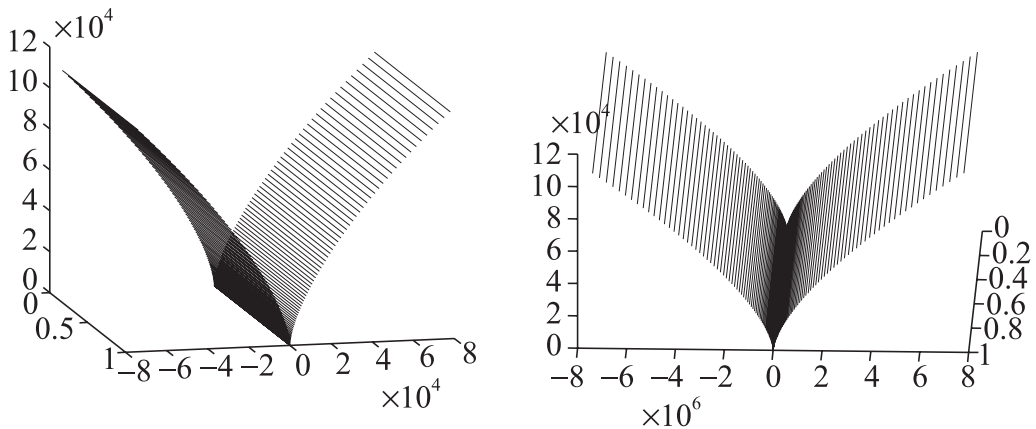


Figure 9.17. The reconstructed developable from two viewing angles.

6. Prove that if a developable σ is represented by two curves $\bar{\sigma}_0, \bar{\sigma}_1$, one the reflection of the other about a line $x = a$, then the representing curves of the rulings are two vertical lines with $\overline{r}r_{23}$ at $x = a$ and $\overline{r}r_{12}$ at $x = 3 - 2a$. This is corollary 9.4.4.
7. Write and implement the algorithm whose input is the representation of $\bar{\sigma}_{123}, \bar{\sigma}_{1'23}$, as a function of y and partitioned in monotone pairs, and outputs the plots of the loci $\overline{r}r_{23}, \overline{r}r_{12}$. This is a good *course project*.

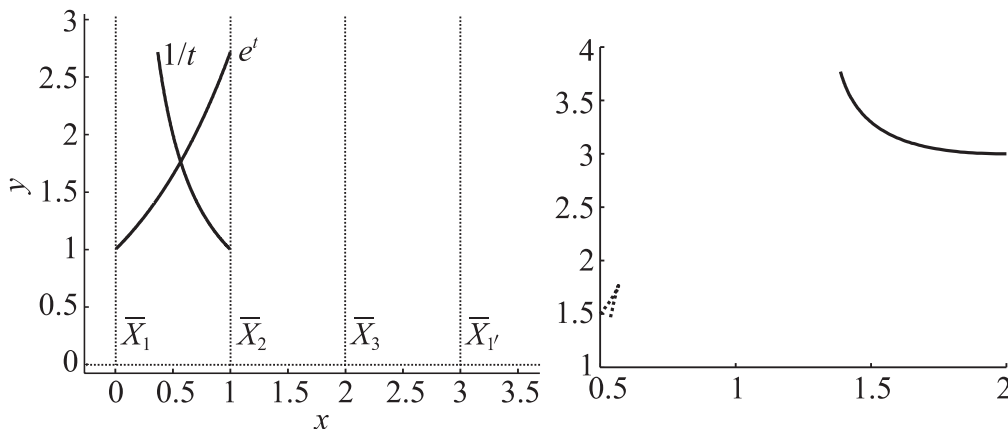


Figure 9.18. Reconstruction of developable from two exponential curves representing it. On the right are the rulings' representing curves: solid \overline{FF}_{12} and dashed lines \overline{FF}_{23} respectively.

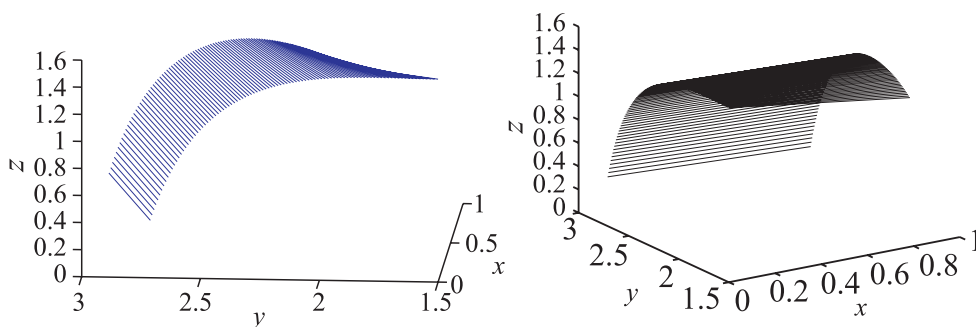


Figure 9.19. The reconstructed developable from two viewing angles.

8. Write and implement the algorithm whose input is the two curves $\overline{FF}_{23}, \overline{FF}_{12}$ (above) and that draws the rulings of the corresponding developable. This is a good *course project*.

9.4.3 Specific Developables

We look now at specific developables, starting with the circular cylinders and cones, and study their representation, rulings, interior points, symmetry properties, and generalizations. Then we continue with more general developables.

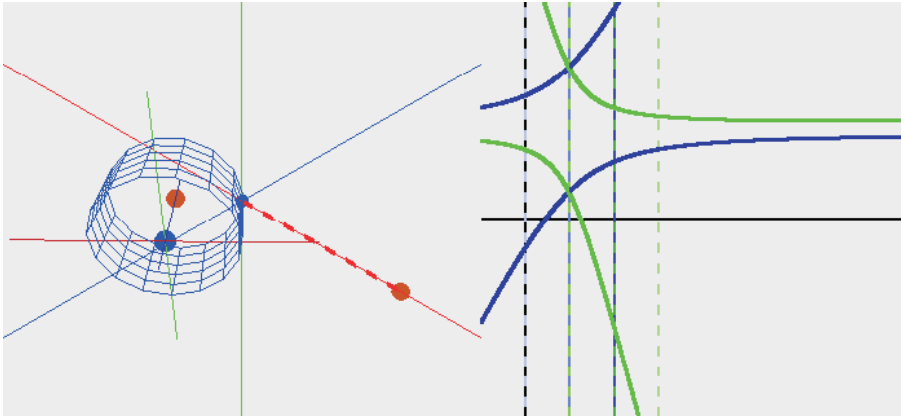


Figure 9.20. ♣ FT-5 A pair of hyperbolas representing the cylinder shown on the left. The • are the “handles” for changing orientation in the software used and have no significance here.

Cylinders

Corollary 9.4.5 (C.K. Hung; Cylinders) *Elliptic cylinders in \mathbb{R}^3 are represented by a pair of hyperbolas.*

Being quadrics by Corollary 9.3.5, they are represented by a pair of conic curves. The cylinder’s orientation determines the number of tangent planes it has parallel to the line u with direction unit vector $\mathbf{u} = 1/\sqrt{3}(\hat{\mathbf{e}}_1 + \hat{\mathbf{e}}_2 + \hat{\mathbf{e}}_3)$, which, in turn, by Corollary 9.3.3 determines the representing curve’s ideal points and hence the type of conics representing the cylinder. For every plane tangent to a circular cylinder there is another tangent plane parallel to it at a distance of D , the cylinder’s diameter, yielding a pair of hyperbolas. So parabolas cannot be representing curves of a circular cylinder. This observation is originally due to A. Chatterjee [93]. Further, the representing curves cannot be ellipses that, according to the section on conic transforms in Chapter 7 on curves, are the images of hyperbolas. The intersection of a plane (including the superplanes) with a cylinder is an ellipse or a pair of lines in the degenerate case. A circular cylinder is easily transformed to one having elliptical sections perpendicular to its symmetry axis, and therefore these observations are equally true for elliptical cylinders.

A cylinder’s rulings are parallel; hence its $\overline{r}r_{12}$ and $\overline{r}r_{23}$ must be two vertical segments. Figures 9.20 and 9.21 are two examples of the cylinders’ representation by hyperbolas. As noted earlier (Exercise 5), the points of intersection of $\bar{\sigma}_{0'} \cap \bar{\sigma}_{1'}$ represent the \bar{r}_{23} of a ruling. Hence the interval between the two points of intersection is the $\overline{r}r_{23}$ of the cylinder’s rulings. This shows an interesting property of hyperbolic curves. Namely, at any point of $\overline{r}r_{23}$ there are two lines tangent to hyperbolas (above

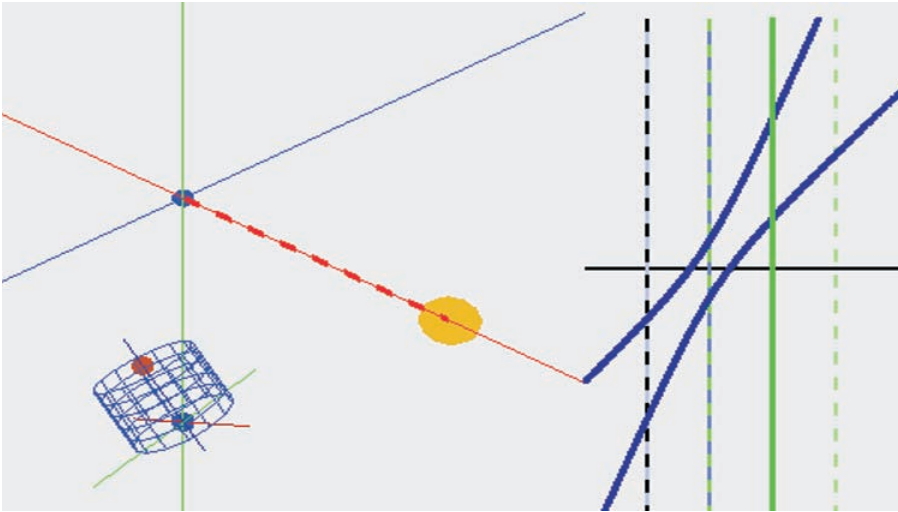


Figure 9.21. Two hyperbolas, one coincident with the \bar{X}_3 axis, representing the cylinder.

or below) at points with the same y -coordinate, since these are the representing points of a tangent plane (Exercise 2 below).

Let us look at this more carefully with the aid of Fig. 9.22 (left), starting with the top point of $\bar{r}\bar{r}_{23}$, the vertical interval between the two intersections. The construction is shown for determining the top point of $\bar{r}\bar{r}_{12}$ by converting the straight-line tangent on $\bar{\sigma}_{1'23} = \bar{\sigma}_{1'}$ to a polygonal line in the $\bar{X}_1, \bar{X}_2, \bar{X}_3$ axis system and intersecting it with the straight line tangent to $\bar{\sigma}_{123} = \bar{\sigma}_0$, and similarly for the bottom point, and thus all of $\bar{r}\bar{r}_{12}$ is determined. If needed, the interval $\bar{r}\bar{r}_{13}$ is found in the same way. By the way, once the representation of a cylinder's ruling is found, the cylinder's orientation, being the same, is also known.

Next, the pair of points representing a specific ruling need to be matched, as shown in Fig. 9.22 (right). Choosing a point $\bar{r}_{12} \in \bar{r}\bar{r}_{12}$, the tangents \bar{T}, \bar{T}' from it to the two branches of $\bar{\sigma}_0$ intersect the line $\bar{r}\bar{r}_{23}$ at two points $\bar{r}_{23}, \bar{r}'_{23}$ corresponding to the two r, r' rulings with the same \bar{r}_{12} . The points of tangency are the representing points $\bar{\pi}_0, \bar{\pi}'_0$ of two tangent planes. The points $\bar{r}_{12}, \bar{r}_{23}, \bar{\pi}_0$ are on the tangent line \bar{T} , and it is easy to see that also $\bar{r}_{1'2}, \bar{r}_{23}, \bar{\pi}_{1'}$ are collinear (not shown). We can conclude that the line r is on the tangent plane π and hence is a ruling, and similarly for r' . Hence T, T' are points on r, r' , and in fact, on the cylinder's intersection with π_0^s , the first sp forming the curve σ_0 . Any point $\bar{\ell}_{23}$ between $\bar{r}_{23}, \bar{r}'_{23}$ together with \bar{r}_{12} represents a line ℓ parallel to the rulings but *interior* to the cylinder. So any point on ℓ is in the cylinder's interior. Conversely, to check whether a point P is

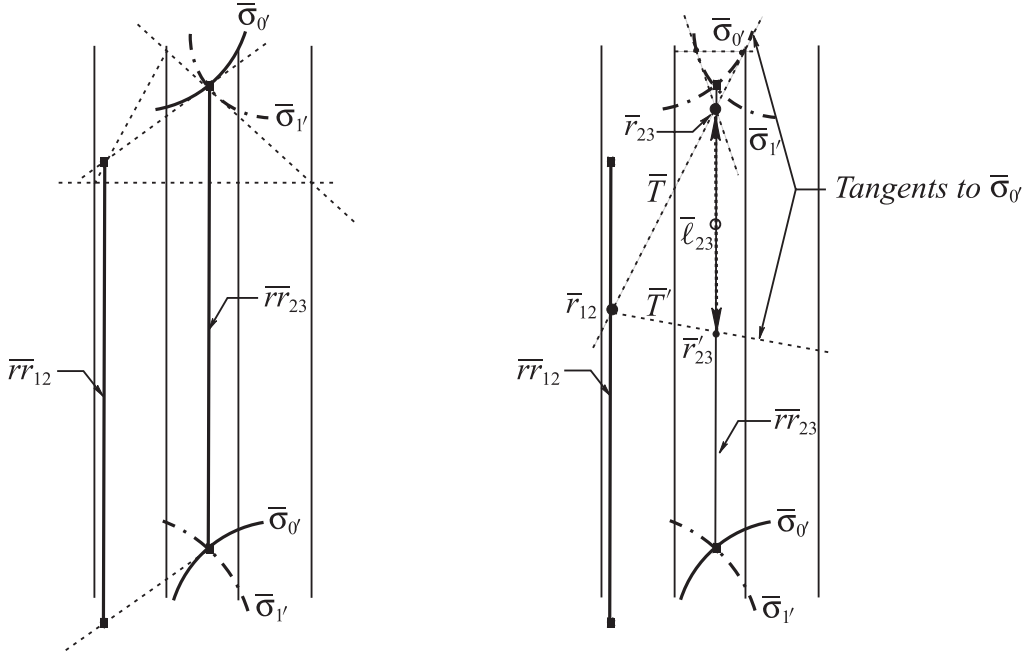


Figure 9.22. Portions of two hyperbolas $\bar{\sigma}_{0'} = \bar{\sigma}_{123}$, $\bar{\sigma}_{1'} = \bar{\sigma}_{1'23}$ representing a cylinder.

(Left) The vertical interval between the two intersections, which necessarily have the same x -coordinate, is the locus $\bar{r}\bar{r}_{23}$ of the cylinder's rulings. The top point of the vertical interval $\bar{r}\bar{r}_{12}$ is constructed from the top intersection, and the lowest point is at the intersection of the vertical line and the tangent to $\bar{\sigma}_{0'}$ at the low intersection point. (Right) For each $\bar{r}_{12} \in \bar{r}\bar{r}_{12}$ there are two rulings r, r' . The $\bar{r}_{23}, \bar{r}'_{23}$ are at the intersections of the tangents to $\bar{\sigma}_{0'}$ from $\bar{r}_{12} \in \bar{r}\bar{r}_{12}$. Between these two points, any other point $\bar{\ell}_{23} \in \bar{r}\bar{r}_{23}$ together with \bar{r}_{12} represents a line ℓ internal to the cylinder parallel to its rulings.

interior, it suffices to check whether there is an interior line ℓ parallel to the rulings with $P \in \ell$ (Exercise 5).

**** A cylinder's representation as a function of its orientation.** The examination of the dependence of a cylinder's representation on its orientation may be safely omitted (up to and including (9.59)) by readers not interested in these details.

To obtain the equation of a circular cylinder with general orientation, we start with the cylinder studied earlier,

$$CY: x_2^2 + x_3^2 = r^2, \quad (9.46)$$

which is symmetric about the x_1 axis, with radius r . The rotation of the coordinate system about the origin is applied using the transformation derived in Section 2.8

and is restated for convenience. The coordinate system x_1, x_2, x_3 is rotated about the origin O to the system x'_1, x'_2, x'_3 , all points remaining stationary, so that a point P has two sets of coordinates $(x_1, x_2, x_3), (x'_1, x'_2, x'_3)$. Letting the direction cosines of the x'_i axes with respect to x_i be $\cos \alpha_i, \cos \beta_i, \cos \gamma_i$ for $i = 1, 2, 3$, respectively, we obtain

$$\begin{aligned} \cos^2 \alpha_i + \cos^2 \beta_i + \cos^2 \gamma_i &= 1, \cos \alpha_j \cos \alpha_k + \cos \beta_j \cos \beta_k + \cos \gamma_j \cos \gamma_k, \\ j &\neq k, i, j, k = 1, 2, 3. \end{aligned} \quad (9.47)$$

The first is due to the property of direction cosines, and the second due to the orthogonality between all pairs of axes $x_j, x_k, i \neq j \neq k$. The transformation relating the two coordinate systems is

$$\begin{aligned} (x'_1, x'_2, x'_3, 1) &= (x_1, x_2, x_3, 1) \begin{pmatrix} \cos \alpha_1 & \cos \beta_1 & \cos \gamma_1 & 0 \\ \cos \alpha_2 & \cos \beta_2 & \cos \gamma_2 & 0 \\ \cos \alpha_3 & \cos \beta_3 & \cos \gamma_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= (x_1, x_2, x_3, 1)A, \end{aligned} \quad (9.48)$$

$$\begin{aligned} (x_1, x_2, x_3, 1) &= (x'_1, x'_2, x'_3, 1) \begin{pmatrix} \cos \alpha_1 & \cos \alpha_2 & \cos \alpha_3 & 0 \\ \cos \beta_1 & \cos \beta_2 & \cos \beta_3 & 0 \\ \cos \gamma_1 & \cos \gamma_2 & \cos \gamma_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= (x'_1, x'_2, x'_3, 1)A^T. \end{aligned} \quad (9.49)$$

The matrix A is orthogonal with determinant $|A| = 1$ and inverse its transpose A^T , which is the matrix of the inverse transformation $\{x'_1, x'_2, x'_3\} \rightarrow \{x_1, x_2, x_3\}$.

To apply the rotation transformation, CY is expressed in matrix form:

$$\begin{aligned} CY: (x_1, x_2, x_3, 1) &\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -r^2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} \\ &= (x_1, x_2, x_3, 1)C(x_1, x_2, x_3, 1)^T. \end{aligned} \quad (9.50)$$

Applying the transformation yields the cylinder

$$CG: (x'_1, x'_2, x'_3, 1)A^T C A(x'_1, x'_2, x'_3, 1)^T \quad (9.51)$$

with the matrix $A^T CA$ written below:

$$\begin{pmatrix} \cos^2 \alpha_2 + \cos^2 \alpha_3 + & \cos \alpha_2 \cos \beta_2 + \alpha_3 \cos \beta_3 \\ \cos \alpha_2 \cos \beta_2 + \alpha_3 \cos \beta_3 & \cos^2 \beta_2 + \cos^2 \gamma_2 \\ \cos \alpha_2 \cos \gamma_2 + \cos \alpha_3 \cos \gamma_3 & \cos \beta_2 \cos \gamma_2 + \cos \beta_3 \cos \gamma_3 \\ 0 & 0 \\ \cos \alpha_2 \cos \gamma_2 + \cos \alpha_3 \cos \gamma_3 & 0 \\ \cos \beta_2 \cos \gamma_2 + \cos \beta_3 \cos \gamma_3 & 0 \\ \cos^2 \gamma_2 + \cos^2 \gamma_3 & 0 \\ 0 & -r^2 \end{pmatrix}$$

Applying the conditions of (9.47) results in

$$A^T CA = \begin{pmatrix} 1 - \cos^2 \alpha_1 & -\cos \alpha_1 \cos \beta_1 & -\cos \alpha_1 \cos \gamma_1 & 0 \\ -\cos \alpha_1 \cos \beta_1 & 1 - \cos^2 \beta_1 & -\cos \beta_1 \cos \gamma_1 & 0 \\ -\cos \alpha_1 \cos \gamma_1 & -\cos \beta_1 \cos \gamma_1 & 1 - \cos^2 \gamma_1 & 0 \\ 0 & 0 & 0 & -r^2 \end{pmatrix}.$$

Completing the evaluation, with $a_i = \cos \alpha_i$, yields

$$\begin{aligned} CG : \sum_{i=1}^3 (1 + a_i^2) x_i^2 - \{(a_1 x_1 + a_2 x_2)^2 + (a_1 x_1 + a_3 x_3)^2 + (a_2 x_2 + a_3 x_3)^2\} \\ = r^2, \end{aligned} \quad (9.52)$$

where the superscript $'$ is dropped. Note that the cylinder is given only in terms of the α_i , which are the direction angles of its axis of symmetry, which we call s . For example, for $a_1 = 1, a_2 = a_3 = 0$, s is the x_1 axis, $a_2 = 1, a_1 = a_3 = 0$, s is the x_2 axis. For $a_1 = a_2 = a_3 = 1/\sqrt{3}$, the axis of symmetry s is the line earlier called u having direction the unit vector $\hat{\mathbf{u}} = 1/\sqrt{3}(\hat{e}_1 + \hat{e}_2 + \hat{e}_3)$, or explicitly, yielding

$$CG : x_1^2 + x_2^2 + x_3^2 - (x_1 x_2 + x_1 x_3 + x_2 x_3) = \frac{3}{2} r^2. \quad (9.53)$$

From Lemma 9.3.1, the representation of CG is the image of two curves c^0, c^1 , which are its intersections with the first two superplanes (see discussion of previous example (9.17) to (9.28)). Proceeding, for the curve $(\overline{CG})_{12} = \overline{CG} \cap \pi_{0'}^s = \bar{c}^0$ it suffices to find $(\overline{CG} \cap \pi_{0'}^s)_{23} = \bar{c}_{23}^0$. With $\pi_{0'}^s : x_1 - 2x_2 + x_3 = 0$, substituting $x_1 = 2x_2 - x_3$ into (9.52) yields the c^0 curve's projection on the $x_2 x_3$ plane:

$$(CG \cap \pi_{0'}^s)_{23} = c_{23}^0 : (5 - B_1^2)x_2^2 + (2 - B_2^2)x_3^2 - 2(2 + B_1B_2)x_2x_3 = r^2, \quad (9.54)$$

where $B_1 = (2a_1 + a_2)$ and $B_2 = (a_3 - a_1)$. The discriminant (see (9.22)) is

$$\Delta = A_4^2 - A_1A_2 = -6 + B_1^2 + B_2^2 + (B_1 + 2B_2)^2,$$

since the intersection of a cylinder with a plane is an ellipse $\Delta < 0$ except possibly for a degenerate case. Recall that the superplanes are the pencils of planes on u such that any superplane, and so also $\pi_{0'}^s, \pi_{1'}^s$, intersects the cylinder given by (9.53) in two parallel lines (rulings), and since $\Delta = 0$ in this case, they are considered a degenerate parabola. Otherwise, the image $\overline{CG}_{23} = \overline{(CG \cap \pi_{0'}^s)}_{23}$ is always a hyperbola and can be found via (9.22) using the interaxis distance $d = 1$ and afterward translating x by 1 (the distance of the \bar{X}_2 from the y axis). It can also be computed without the translation of x from the image of $(CG \cap \pi_{0'}^s)_{12} = c_{12}^0$ if available, and in either case the result is the same:

$$\overline{(CG)}_{12} = \overline{CG}_{0'} : Ix^2 + (\Delta/r^2)y^2 - [3 - B_2(B_1 + B_2)]x - (2 - B_2^2) = 0, \quad (9.55)$$

where $I = 3 - (a_1 + a_2 + a_3)^2$. As mentioned, the cylinder's axis of symmetry passes through the origin. Translating the origin does not change the coefficients of the quadratic terms in (9.65) and therefore has no effect in determining the type of conic obtained for the image $\overline{(CG)}_{12}$.

In order to make a comparison later we also find the image $\overline{(CG)}_{23}$. Proceeding as before for the intersection $\overline{(CG)}_{23} = \overline{CG \cap \pi_{1'}^s} = \bar{c}^1$, it suffices to find $\overline{(CG \cap \pi_{1'}^s)}_{23} = \bar{c}_{23}^1$ with $\pi_{1'}^s : x_1 + 2x_2 - x_3 = 0$, substituting $x_1 = 2x_3 - x_2$ into (9.65) to obtain the c^1 curve's projection on the x_2x_3 -plane:

$$(CG \cap \pi_{1'}^s)_{23} = c_{23}^1 : (2 - C_1^2)x_2^2 + (5 - C_2^2)x_3^2 - 2(2 + C_1C_2)x_2x_3 = r^2, \quad (9.56)$$

where $C_1 = (a_2 - a_1)$ and $C_2 = (2a_1 + a_3)$. Its transform is

$$\begin{aligned} \overline{CG}_{23} = \overline{CG}_{1'} : I(x+1)^2 + (\Delta/r^2)y^2 - [3 - C_2(C_1 + C_2)](x+1) - (2 - C_2^2) \\ = 0. \end{aligned} \quad (9.57)$$

The translation $(x+1)$ is due to the \bar{X}_2 axis being one unit from the y axis. This together with (9.54) describes the pair of hyperbolas in the representation. Returning to the representation of the cylinder CD given by (9.53) it is found from (9.55) by

substituting $a_1 = a_2 = a_3 = 1/\sqrt{3}$, together with $x_1 = 2x_2 - x_3$, and is the pair of lines

$$\begin{cases} x_3 = x_2 \pm r/\sqrt{2}, \\ x_2 = x_1 \pm r/\sqrt{2}. \end{cases} \quad (9.58)$$

Each of these lines is represented by two vertical lines. So the degenerate cylinder representation consists of the four vertical lines $\bar{\ell}_{12}, \bar{\ell}_{23}$, two for each index pair:

$$\bar{\ell}_{12} : x = \pm r/\sqrt{2}, \quad \bar{\ell}_{23} : x = 1 \pm r/\sqrt{2}. \quad (9.59)$$

In vector notation, a circular cylinder is given by

$$\mathbf{x}(t, v) = \mathbf{b} + r(\hat{\mathbf{s}}_1 \cos t + \hat{\mathbf{s}}_2 \sin t) + v\hat{\mathbf{s}}, \quad (9.60)$$

where $\hat{\mathbf{s}}$ is the unit vector in the axis direction and the two other unit vectors are such that $\mathbf{s} = \hat{\mathbf{s}}_1 \times \hat{\mathbf{s}}_2$, which are implicitly given in terms of the basic unit vectors $\hat{\mathbf{e}}_i, i = 1, 2, 3$. The two parameters v, t are real numbers with $0 \leq t \leq 2\pi$. The constant vector \mathbf{b} and the cylinder's axis of symmetry s share a point, which is the center of a circle of radius r , and a point on its circumference is given by the vector $\mathbf{p} = r(\hat{\mathbf{s}}_1 \cos t + \hat{\mathbf{s}}_2 \sin t)$. This vector is normal to the tangent plane

$$\pi(t) : (\hat{\mathbf{s}}_1 \cos t + \hat{\mathbf{s}}_2 \sin t) \cdot \mathbf{x} = (\hat{\mathbf{s}}_1 \cos t + \hat{\mathbf{s}}_2 \sin t) \cdot \mathbf{b} + r \quad (9.61)$$

containing the ruling parallel to \mathbf{s} given by a fixed value of t in (9.60). The representation can be found directly from Theorem 9.4.1 and the family of tangent planes above as $t \in [0, 2\pi]$ rather than from the intersections with the superplanes π_0^s, π_1^s . To do this, (9.61) is rewritten as

$$\pi(t) : \sum_{i=1}^3 (A'_i \sin t + B'_i \cos t)x_i = (A_0 \sin t + B_0 \cos t + C_0), \quad (9.62)$$

where the constants A'_i, B'_i, C'_i derive from the decomposition of the unit vectors $\hat{\mathbf{s}}, \hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2$ in terms of the $\hat{\mathbf{e}}_i, i = 1, 2, 3$. The parallel-coordinate representation of the planes $\pi(t)$ are the points with homogeneous coordinates $(x'(t), y'(t), S(t))$:

$$\begin{cases} x'(t) &= A_1 \sin t + B_1 \cos t, \\ y'(t) &= A_0 \sin t + B_0 \cos t + C_0, \\ S(t) &= A_4 \sin t + B_4 \cos t, \end{cases} \quad (9.63)$$

where $S(t)$ is the sum of the x_i coefficients. As t varies, a curve on the projective plane is traced having ideal points where $S(t) = 0$. The detailed form of the

constants is immaterial to the subsequent calculation. Solving for $\cos t$ and $\sin t$ and substituting in the trigonometric identity $\sin^2 t + \cos^2 t = 1$ yields

$$\left(\frac{A_4 x' - A_1 S}{A_4 B_1 - A_1 B_4} \right)^2 + \left(\frac{B_1(y' - C_0) - B_0 x'}{A_1 B_0 - A_0 B_1} \right)^2 = 1. \quad (9.64)$$

With the replacements $x = x'/S$, $y = y'/S$, this equation describes a regular or degenerate conic, in this case, one of the representing hyperbolic curves of the cylinder in (9.60).

Generalized cylinders. A surface's description in vector form is convenient and also clarifies its structure. It was proposed by the mathematical physicist W.J. Gibbs, who was also an early advocate for *visualization* in science⁴⁵ [68], [69].

A *general cylinder* is specified by the equation

$$\mathbf{x}(s, v) = \mathbf{c}(s) + v\mathbf{r}, \quad (9.65)$$

where $\mathbf{c}(s)$ is a space curve given in terms of its arc length s , v is the second parameter, and \mathbf{r} a constant vector. Its rulings are given by $v\mathbf{r}$ with specific values of s and are therefore parallel. In (9.60), $\mathbf{c}(s)$ is the circle with radius r centered at \mathbf{b} , and the rulings are $v\mathbf{s}$. We can conclude that the intersections of $\pi_{0'}^s, \pi_{1'}^s$ are perspectives (in the sense of perspective transformations in projective geometry) of curves whose images provide the representing curves. By the way, this form is dimension-invariant, and we can consider it as defining a surface in \mathbb{R}^N , a developable being a hypersurface with a one-parameter family of tangent planes represented by $N - 1$ curves. From now on, unless otherwise specified, cylinder refers to a general cylinder. In Fig. 9.23 a cylinder is shown having cusps along one of its rulings. Its two representing curves, each with an inflection point, remind us of the known *developable* \leftrightarrow *curve* duality. In fact, we met an example of this duality earlier, in Fig. 9.17, also showing a generalized cylinder having two representing curves with inflection points, Fig. 9.16, with $\overline{r}r_{12}, \overline{r}r_{23}$ two vertical lines due to its parallel rulings. There is a further pleasing duality here: the representing curves seen in Fig. 9.24 indicate that a plane tangent to *two* rulings (i.e., *bitangent*) exists and is represented by crossing points, one on each curve, akin to the *bitangent* \leftrightarrow *crossing point* for curves. From our previous discussions, it is clear that the vertical semi-infinite line $\overline{r}r_{23}$ goes through the crossing point that after all, is the common \bar{r}_{23} of the two rulings on the *bitangent* plane. This also points out that the crossing point and cusps in the representing curves seen earlier in Figs. 9.9 and 9.10 correspond to bitangent planes and rulings that are the loci of inflection points in the developables.

⁴⁵For the importance of visual thinking in science see [122].

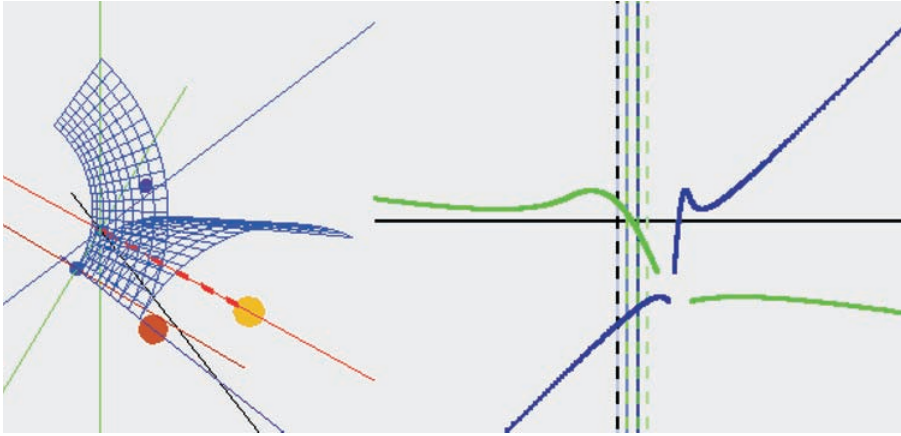


Figure 9.23. ♣ **FT-6** A general cylinder illustrating the *developable* \leftrightarrow *curve* duality. The ruling formed by cusps is transformed to an *inflectio* point in each of the representing curves.

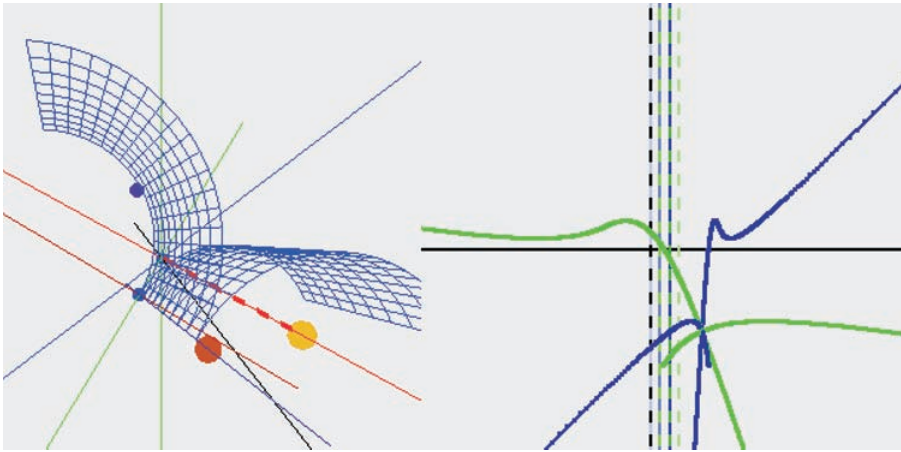


Figure 9.24. ♣ **FT-6e** The two leaves of the surface in the previous figure are extended. Here the surface, in addition to the line of cusps, has a *bitangent plane*, which is represented by the two representing curves *crossing at the same point*. Such a hypersurface in \mathbb{R}^N is represented by $N - 1$ curves with crossing points and inflection points at the same y coordinate (Exercise 16).

Symmetry and coincidence. A symmetric pair of representing curves necessarily represents a cylinder, as stated in Corollary 9.4.4. An example of the developable's reconstruction from its rulings is shown in Fig. 9.16. Earlier, we also saw that a cylinder with (9.17) having rulings parallel to the x_1 axis has two coincident

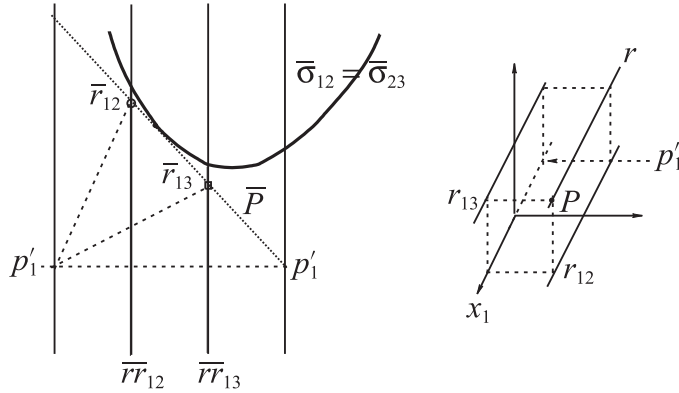


Figure 9.25. Coinciding curves represent a cylinder with rulings parallel to the x_1 axis. The $\bar{r}\bar{r}_{12}$ is on the \bar{X}_2 axis and $\bar{r}\bar{r}_{13}$ on the \bar{X}_3 axis. A ruling r on a point P and its representation in \parallel -coords is shown.

hyperbolic representing curves. These are instances of general results characteristic of cylinders, which, together with variants of Corollary 9.4.4, were discovered by C.K. Hung [93].

Corollary 9.4.6 (C.K. Hung; Coincident representing curves) A developable σ in \mathbb{R}^3 is a cylinder with rulings parallel to the x_1 axis if and only its two representing curves $\bar{\sigma}_{12}$, $\bar{\sigma}_{23}$ coincide.

Proof. An instance of the two curves $\bar{\sigma}_{12}$, $\bar{\sigma}_{23}$ coinciding, shown in Fig. 9.25, clarifies the proof. A tangent \bar{P} at any point represents a point on σ_0' with coordinates (p_1, p_2, p_3) and also a point on σ_1' with coordinates $P' = (p_1', p_2, p_3)$. With respect to the axes \bar{X}_1' , \bar{X}_2 , \bar{X}_3 , \bar{P}' is a straight line (coincident with \bar{P}) and a polygonal line with respect to \bar{X}_1 , \bar{X}_2 , \bar{X}_3 . The intersection $\bar{P} \cap \bar{P}'$ contains the interval of \bar{P} between the \bar{X}_2 and \bar{X}_3 axes. This shows that the line r on P , P' is perpendicular to the x_2x_3 plane at the point $(0, p_2, p_3)$ and therefore parallel to the x_1 axis. Note that \bar{r}_{12} is on the \bar{X}_2 axis at the value p_2 . Since this is true for *any* point, the \bar{r}_{12} are on a vertical line $\bar{r}\bar{r}_{12}$ on the \bar{X}_2 axis. For the second point in representation of the r we can use the \bar{r}_{13} , which are all on a vertical line $\bar{r}\bar{r}_{13}$ on the \bar{X}_3 axis. So the lines r formed by the points P , P' are all parallel, so they are rulings and σ must be a cylinder. The $\bar{r}\bar{r}_{12}$, $\bar{r}\bar{r}_{13}$ may be complete or semi-infinite lines or finite intervals.

For the proof in the opposite direction assume that σ is a cylinder with rulings parallel to the x_1 axis. Then $\bar{r}\bar{r}_{12}$, $\bar{r}\bar{r}_{13}$ are the \bar{X}_2 , \bar{X}_3 axes respectively. Choose a pair \bar{r}_{12} , \bar{r}_{13} representing a ruling r . The points are the intersections of a line \bar{P} tangent to $\bar{\sigma}_0'$ and another line \bar{Q} tangent to $\bar{\sigma}_1'$. Because \bar{r}_{12} , $\bar{r}_{13} \in \bar{P} \cap \bar{Q}$,

the whole *interval* from \bar{r}_{12} to \bar{r}_{13} is contained in $\bar{P} \cap \bar{Q}$. Two lines sharing an interval must be identical, that is, $\bar{P} = \bar{Q}$. So these must be tangents to the same curve possibly at two *different* points, which must have the *same* y value. The curve representing a developable is locally monotonic; hence the points of tangency must be identical if their y values are equal. We can conclude that the two representing curves coincide. ■

FT-7 Cones

A developable all of whose rulings intersect on a single *characteristic point*, recall (9.31), is a *cone* whose representing curves and rulings are studied next.

Corollary 9.4.7 (Cones; C.K. Hung) *Circular Cones in \mathbb{R}^3 are represented by a pair of conic curves.*

Circular cones, being quadrics, are, by Corollary 9.3.5, represented by a pair of conic curves whose type is determined, via Corollary 9.3.3, by the number of tangent planes the cone has parallel to the line u in the direction of the unit vector $\hat{u} = 1/\sqrt{3}(\hat{e}_1 + \hat{e}_2 + \hat{e}_3)$. Unlike cylinders, all three types of conics can represent curves, as stated in Table 9.1. The angle θ between the cone's ruling and the axis of symmetry s is constant. So for \hat{u} to be parallel to π_t , one of the cone's tangent planes, there must exist a plane π containing \hat{u} such that the angle $\pi \cap s$ equals θ . Consider two tangent planes π_{t_1}, π_{t_2} intersecting at the line ℓ . If u is parallel to ℓ , then it is also parallel to the two tangent planes, and it is clear that u cannot be parallel to more than two tangent planes.

Concerning the rulings from Figs. 9.27, 9.29, 9.31 where one of conics collapses to a line or line segment, we learn that at least one of the curves $\overline{rr}_{23}, \overline{rr}_{12}$ representing the cone's ruling is that line (or segment). Further discussion of the rulings, interior lines, and points of a cone are left to as Exercises 10 and 11. By the way, since $\overline{rr}_{12}, \overline{rr}_{23}$ cannot be vertical, as for cylinders the hyperbolas, representing cones differ from those representing cylinders (Exercise 13).

Table 9.1. Cones are represented by conic curves.

Number of tangent planes parallel to u	Represented by a pair of
0	ellipses
1	parabolas
2	hyperbolas

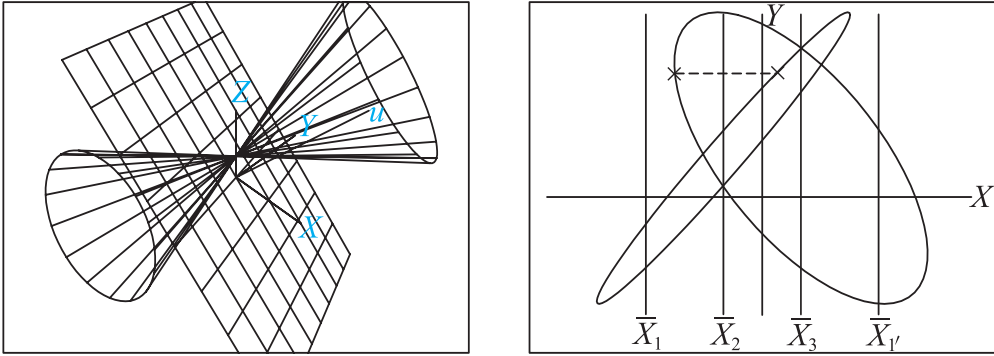


Figure 9.26. A circular (double) cone without tangent planes parallel to the line u . The cone is represented by two ellipses. Two points, one on each ellipse, represent one of the tangent planes.



Figure 9.27. Another (double) cone without tangent planes parallel to u . It is represented by two ellipses, one collapsing to a line segment. This is the representation of the *double cone*, though here and subsequently only one cone is shown.

**** The representation of cones as a function of the orientation and apex angle.**

As for cylinders, we study a right circular cone CC with vertex at \mathbf{a} from its equation obtained from the observation [128, p. 32] that any ruling has the same angle θ , with the symmetry axis having direction the unit vector $\hat{\mathbf{s}} = s_1\hat{\mathbf{e}}_1 + s_2\hat{\mathbf{e}}_2 + s_3\hat{\mathbf{e}}_3$. For $\mathbf{x} \in CC$ and $k = \cos \theta$, the angle between $\mathbf{x} - \mathbf{a}$ (a ruling on CC) and $\hat{\mathbf{s}}$ is θ or $\pi - \theta$, letting

$$CC : |(\mathbf{x} - \mathbf{a}) \cdot \hat{\mathbf{s}}| = k|\mathbf{x} - \mathbf{a}|.$$

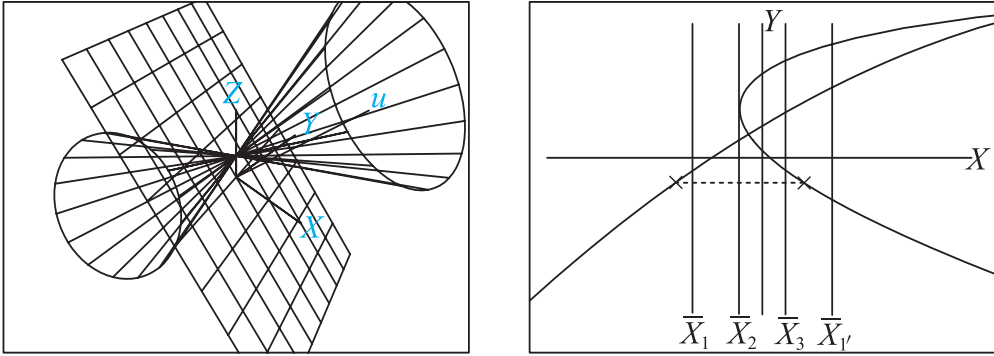


Figure 9.28. A circular (double) cone with one tangent plane parallel to the line u . The cone is represented by two parabolas. Two points, one on each parabola, represent one of the tangent planes.

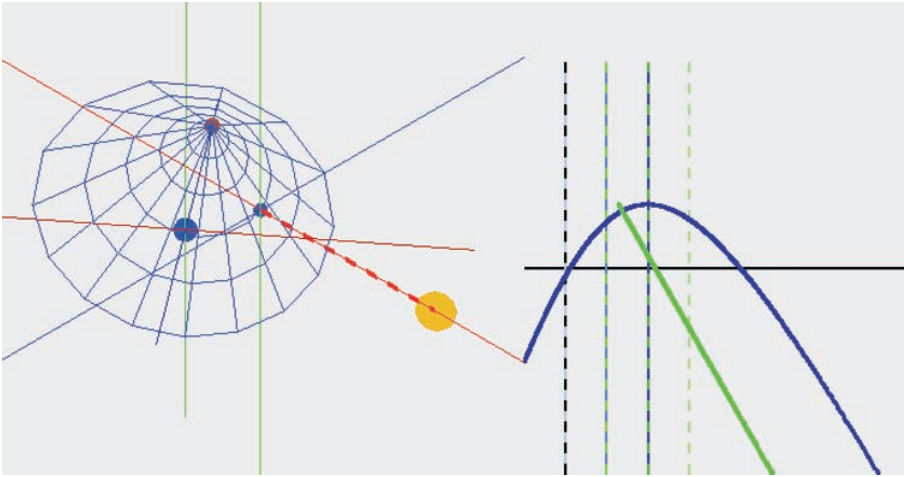


Figure 9.29. Another circular (double) cone with one tangent plane parallel to the line u . The cone is represented by two parabolas, one collapsing to a half-line.

It is convenient to square both sides and write the cone's equation as

$$CC : [(\mathbf{x} - \mathbf{a}) \cdot \hat{\mathbf{s}}]^2 = k^2 (\mathbf{x} - \mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}). \quad (9.66)$$

For the sake of simplicity, the cone's vertex is placed at the origin $(0, 0, 0)$, for the translation to \mathbf{a} does not change the type of representing conic curve. We find the

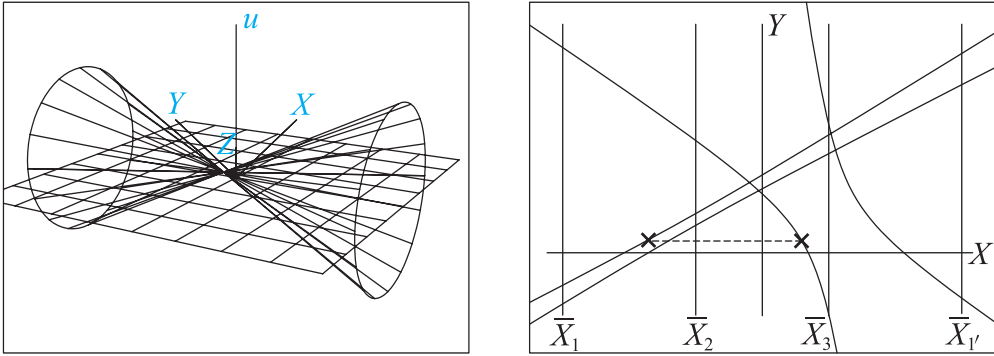


Figure 9.30. A circular cone with two tangent planes parallel to the line u . The cone is represented by two hyperbolas. The two points, one on each hyperbola, represent one of the tangent planes.

intersection with the first superplane π_0^s , in fact, the projection on the x_2x_3 plane, namely

$$(CC \cap \pi_0^s)_{23} = c_{23}^0 : (-5k^2 + D^2)x_2^2 + (-2k^2 + C^2)x_3^2 + (2k^2 - CD)x_2x_3 = 0, \quad (9.67)$$

where $C = (s_1 - s_3)$, $D = (2s_1 + s_2)$. We recall that the signs of the indicator I and discriminant Δ determine the type of conic curve. They are

$$\begin{aligned} I &= -3k^2 + (s_1 + s_2 + s_3)^2, \quad \Delta = -6k^4 + k^2[(2C - D)^2 + C^2 + D^2] \\ &= -6k^4 + k^2[3(s_1 - s_3)^2 + 2], \end{aligned}$$

using $\sum_{i=1}^3 s_i^2 = 1$, $\sum_{j \neq k} s_j s_k = 0$ in the last step:

$$I = 0 \Rightarrow k = \pm \frac{|s_1 + s_2 + s_3|}{\sqrt{3}}, \quad \Delta = 0 \Rightarrow k = \pm \sqrt{\frac{(3(s_1 - s_3)^2 + 2)}{6}}. \quad (9.68)$$

Elliptic representing curves, Figs. 9.26, 9.27, are produced when $I < 0$, $0 < \Delta$ and there are no tangent planes parallel to \mathbf{u} . The vertex angle constraints are $\sqrt{[3(s_1 - s_3)^2 + 2]/6} > |k| > |s_1 + s_2 + s_3|/\sqrt{3}$.

There exist an orientation and a cone, specified by k , simultaneously for $I = 0 = \Delta$ for

$$2(s_1 + s_2 + s_3)^2 = (2s_3 + s_2)^2 + (s_1 - s_3)^2 + (2s_1 + s_2)^2 \Rightarrow s_1 = s_3 = s,$$

$$|k| = \frac{1}{\sqrt{3}}, \quad (9.69)$$

the value of k obtained from the second condition (for $\Delta = 0$) in (9.68). From $\sum_{i=1}^3 s_i^2 = 1$, $s_2 = \sqrt{1 - 2s^2}$, and the first condition in (9.68), $s = s_1 = s_3 = 0$, $s_2 = 1$ or $s = s_1 = s_3 = 2/3$, $s_2 = \pm 1/\sqrt{3}$ are the orientations when the double cone with $k = 1/\sqrt{3}$ has one tangent plane parallel to the line \mathbf{u} , yielding a pair of parabolic representing curves, Figs. 9.28, 9.29. When $\Delta > 0$ and $I = 0$, the representing conics are also parabolas, and an easy calculation yields the orientation constraint $s_1 \neq s_3$. This implies that $\sqrt{[2(s_1 - s_3)^2 + 2]/6} > |k|$ or $|k| \in (1/3, \sqrt{5/6})$.

When $\Delta < 0$, the curve given by (9.67) is an ellipse, and irrespective of the sign of I , its image is a hyperbola, Figs. 9.30 and 9.31. From $\Delta < 0$, we get $\sqrt{3}(s_1 - s_3)^2 + 2 < \sqrt{6}|k|$, and in turn $|k| \in (1/\sqrt{3}, \sqrt{5/6})$ without further restrictions on the orientation (i.e., the s_i). Also of interest here is the combination $\Delta = 0$, the conic $CC \cap \pi_0^s$ being a parabola and $I \neq 0$, so that $s_1 \neq s_3$, with $|k| > 1/\sqrt{3}$, is the image curve is a hyperbola with one vertical asymptote.

For the combination $\Delta > 0, I > 0$, adding the inequalities yields $\sqrt{[3(s_1 - s_3)^2 + 4]/3} < |k|$; the corresponding cones have two tangent planes parallel to \mathbf{u} , producing a pair of hyperbolas, Figs. 9.30, 9.31, as representing

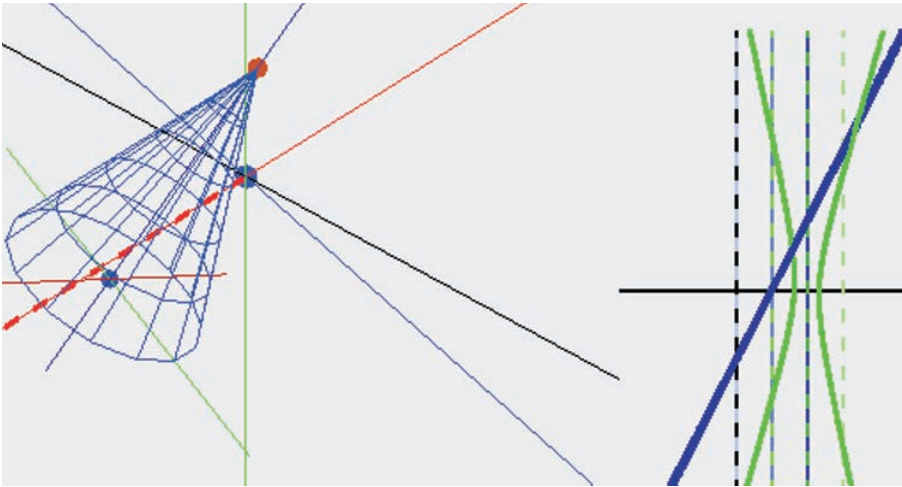


Figure 9.31. Another circular cone with two tangent planes parallel to the line \mathbf{u} . The cone is represented by two hyperbolas, one of them collapsing to a line.

Table 9.2. Summary of representation results for cones.

Pair of curves representing <i>Cone</i>	Orientation & vertex angle $\theta = \cos^{-1} k$
Ellipses – Figs. 9.26, 9.27	$I < 0, \Delta > 0, s_1 \neq s_3$ $\sqrt{[3(s_1 - s_3)^2 + 2]/6} > k > s_1 + s_2 + s_3 /\sqrt{3}$
Parabolas – Figs. 9.28, 9.29	$I = 0, \Delta = 0, k = \pm 1/\sqrt{3}$ $s_1 = s_3 = 0, s_2 = 1,$ $s_1 = s_3 = 2/3, s_2 = \pm 1/\sqrt{3}.$
Parabolas opening vertically	$I = 0, \Delta > 0, s_1 \neq s_3$ $\sqrt{[3(s_1 - s_3)^2 + 2]/6} > k = s_1 + s_2 + s_3 /\sqrt{3}$
Hyperbolas – Figs. 9.30, 9.31	all $I, \Delta < 0, \sqrt{[3(s_1 - s_3)^2 + 2]/6} < k $
Hyperbolas	$I > 0, \Delta > 0$ $\sqrt{[3(s_1 - s_3)^2 + 4]/3} < k $
Hyperbolas with vertical asymptote	$I \neq 0, \Delta = 0, \sqrt{[3(s_1 - s_3)^2 + 2]/6} = k $ $ k \neq s_1 + s_2 + s_3 $

curves. This is also the case for $\Delta = 0, I \neq 0$, constraining the vertex angle by $\sqrt{[3(s_1 - s_3)^2 + 2]/6} = |k|$ and $|k| \neq |s_1 + s_2 + s_3|$; here each hyperbola has one vertical asymptote. These conditions, which are combinations of the cone's size, measured by the vertex angle, and the orientation, producing 0, 1, or 2 tangent planes parallel to $\hat{\mathbf{u}}$, are summarized in Table 9.2. A cone in 5-D whose representing curves are ellipses is shown in Fig. 11.13 in Section 11.3, wherein an innovative approach to surface representation as applied to quadrics is presented. A general cone can be described by

$$\mathbf{x}(s, v) = \mathbf{a} + v\mathbf{t}(s), \quad (9.70)$$

where $\mathbf{t}(s)$ is a unit vector field described on a curve, such as a circle for circular cones, all rulings passing through the vertex at \mathbf{a} and a point of the curve $\mathbf{t}(s)$ [166, pp. 90–91].

9.4.4 ♣ FT-8 General Developable Surfaces

In general, a developable surface can be described by

$$\mathbf{x}(s, v) = \mathbf{y}(s) + v\mathbf{g}(s), \quad |\mathbf{g}(s)| = 1, \quad (\mathbf{y}' + v\mathbf{g}') \times \mathbf{g} = 0. \quad (9.71)$$

The equation describes the more general *ruled* surfaces, and it is the additional condition on \mathbf{y}' , v , \mathbf{g} , \mathbf{g}' that specializes it to developables. An example of an intricate developable is the helicoid

$$x_1 = a \cos s - av \sin s, \quad x_2 = a \sin s + av \cos s, \quad x_3 = b(s + v). \quad (9.72)$$

The representing curves at two orientations, shown in Figs. 9.32, 9.33, are instances of interactive exploration, reminding us of interactivity's importance in visualization. Proving that the helicoid is developable and finding the curves representing its rulings (Exercise 9) are left to the reader. The class of developables consists of planes, cylinders, cones, *tangent surfaces of a curve*, and combinations thereof. Quite a few, e.g., *helicoid*, *polar*, *tangential*, and *rectifying developables*, which are examples of tangent surfaces of a curve [166, pp. 66–72], are difficult to recognize visually as developables. Yet in $\|\cdot\|$ -coords they are immediately recognizable, being represented by curves, and some properties are apparent from the duality *developables* \leftrightarrow *space curves*, i.e., *cusps* \leftrightarrow *inflection-point*, Fig. 9.23; *bitangent plane* \leftrightarrow *crossing-point*, Fig. 9.24. It is significant that for developables in the above examples, the two representing curves are similar. These curves are images

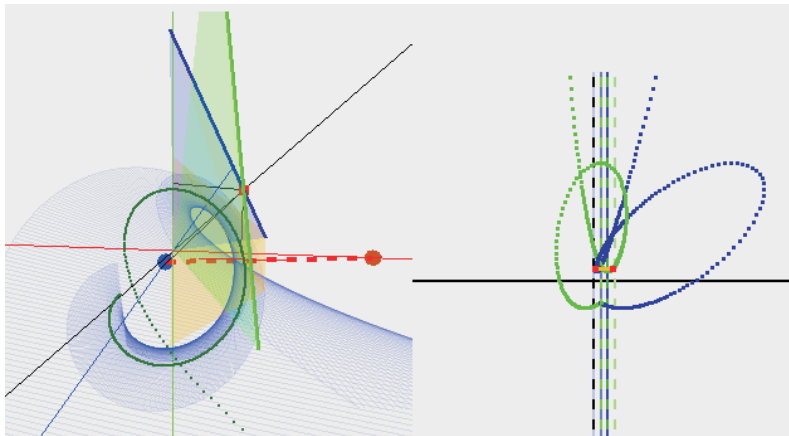


Figure 9.32. Developable helicoid and its two representing curves.

The two points on the right (one on each curve) represent the tangent plane shown on the left determined by the two line intersections with the first $\pi_{O'}^s$ and second $\pi_{I'}^s$ superplanes.

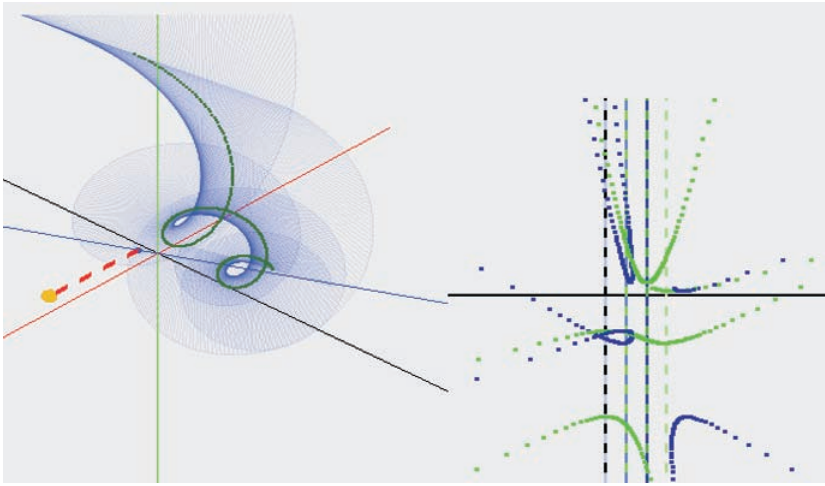


Figure 9.33. At this orientation, the representing curves have several ideal points. They correspond to the helicoid’s tangent planes parallel to \mathbf{u} (making equal angles with all coordinate axes).

of the surfaces’ section with the first two superplanes, so they are projectively equivalent (i.e., one can be obtained from the other by a projective transformation) for the cylinders and cones. For the helicoid, whose rulings are a pencil of lines on its axis, the reason is not so obvious. Still for other developables this may not be true. It would be nice to settle the question whether two very different curves such as those in Fig. 9.10 represent a developable (see Exercise 16). So for developables with similar representing curves, the transition to \mathbb{R}^N is seamless: they are represented by $N - 1$ similar planar curves, properties are revealed by duality, and the reconstruction is analogous, i.e., three linked points on three curves determine each tangent hyperplane for $N = 4$ as discussed in the next section.

In our surroundings there are many objects that look “nearly” developable, and many parts are manufactured by deforming flat sheets. This and the simple $\|\cdot\|$ -coordinate representation motivate the study of “approximate developables” (Exercise 18), as for the proximate planes in Chapter 8, by introducing perturbations in the developables’ equation, allowing $|(\mathbf{y}' + v\mathbf{g}') \times \mathbf{g}| < c \neq 0$ but small (see also the following section on ruled surfaces), and there may be other useful definitions. The conjecture is that families of “nearly developables” are represented in $\|\cdot\|$ -coords by curved “strips” with properties “close” to those discovered by duality (of developables) with representing curves and are amenable to an intuitive treatment. New classes of surfaces (patches) and their N -D generalizations suitable for applications such as geometric modeling may emerge.

9.4.5 ** Higher Dimensions

A developable hypersurface σ in higher dimensions is the envelope of a one-parameter family of hyperplanes. Each of these hyperplanes is one of its tangent hyperplanes. All the results in this section are generalizable to developable hypersurfaces in dimensions $N > 3$. These results are stated without proof below for readers who would like to turn these theories into practical implementations.

Denote the $N - 1$ representing curves by $\bar{\sigma}_j$, where $j \in I$ is one of the indexing sets $\{1, 2, \dots, N\}$, $\{2, 3, \dots, N, 1'\}$, \dots , $\{N - 1, N, 1', 2', \dots, (N - 2)'\}$. If the $\bar{\sigma}_j$'s are monotonic with respect to the y axis, then σ can be reconstructed as follows:

1. Draw a horizontal line to intersect the $\bar{\sigma}_j$'s at $N - 1$ points $\bar{\pi}(t_0)_j$ that together specify a tangent hyperplane $\pi(t_0)$.
2. Draw $N - 1$ tangent lines l_j at each $\bar{\pi}(t_0)_j$ to each $\bar{\sigma}_j$. Each of these lines specifies the intersecting point of the “ruling” (an $(N - 2)$ -flat) of σ at $\pi(t_0)$ with the superplane π_j^s of the corresponding index.
3. Find the N intercepts of each of the $N - 1$ l_j with the corresponding axes to obtain the coordinates of these points of intersection.
4. Draw horizontal lines to bring all numbers to the first set of axes $\{\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N\}$.
5. Follow the construction procedure for 3-D to obtain the $(N - 2)$ -dimensional ruling.

In general, when the representing curves are not monotonic with respect to the y axis, the curves can be partitioned into k traceable pieces at self-tangential points. The representation is unique if one $(N - 1)$ -tuple of points of equal y -value is given for matching one set of pieces, for a total of k such tuples. The reconstruction algorithm still works most of the time by producing a maximum of $O(k^{N-1})$ possible alternative hypersurface *even if the matching tuples are not given a priori*. It fails only when the representing curves have an infinite number of self-tangencies that break the curves into similar pieces, an infinite number of which have the same number of extremal points and the same extremal y -coordinates occurring in exactly the same sequence. Curve dualities have analogues for developables in \mathbb{R}^3 , as illustrated in Figs. 9.23, 9.24 and in higher dimensions.

Exercises

1. Derive the representation and rulings' curves for quadric cylinders (i.e., parabolic and hyperbolic). This is an easy *course project*.

2. For two hyperbolas representing a circular cylinder, prove that for any point on the vertical line $\overline{r}r_{23}$, the tangents to the two hyperbolic branches (above the point) have the same y -coordinate.
3. Prove that if one of the representing curves of a developable has a crossing point, so does the other representing curve. Do crossing points always coincide as in Fig. 9.24?
4. For the representing hyperbolas in Fig. 9.20 construct the vertical intervals:
 - (a) $\overline{r}r_{1'2}$ consisting of the representing points $\bar{r}_{1'2}$ and
 - (b) $\overline{r}r_{13}$ consisting of the representing points \bar{r}_{13} .

Do the same for the representing hyperbolas in Fig. 9.21.

5. Given a circular cylinder CG and a point P , write an algorithm to decide whether P is interior to CG .
6. Generalize Corollary 9.4.6 to 4-D and higher dimensions.
7. What kind of curves can the representing curves of a cone's rulings be?
8. Show that (9.71) defines a developable surface. Hint: find the normal vector and show that it depends on one parameter.
9. Show that the helicoid given by (9.72) is developable and find its representing curves. Hint: check that $(\mathbf{y}' + v\mathbf{g}') \times \mathbf{g} = 0$ as in (9.71).
10. Write and implement an algorithm that inputs a cone's representation and outputs the representing curves $\overline{r}r_{12}, \overline{r}r_{23}$ of its rulings.
11. Give an algorithm that constructs half-lines and points interior to a cone.
12. Find the curves representing the rulings of the helicoid given by (9.72).
13. How do the hyperbolas representing cones differ from those representing cylinders?
14. Generalize the representation of circular cones to N -D (see also Section 11.3). In particular, are there cases in which different conic curves appear simultaneously in the representation of a cone in N -D?
15. Prove that the representing curve σ_0 of a developable σ is composed of the *envelopes* of the family of lines \bar{R} on the points $\bar{r}_{12}, \bar{r}_{13}, \bar{r}_{23}$ of the rulings r of the developable. Similarly, the curve $\sigma_{1'}$ is the envelope of the family of lines \bar{R}' on $\bar{r}_{1'2}, \bar{r}_{1'3}, \bar{r}_{23}$.
16. Prove or give a counterexample that the two representing curves of a developable in \mathbb{R}^3 are projectively equivalent (i.e., one can be obtained from the other via a planar projective transformation) and generalize to \mathbb{R}^N ; this is a *research problem*.
17. Dualities for developables in \mathbb{R}^N (see Figs. 9.23, 9.24). Prove that
 - (a) a line of cusps is represented by N inflection points having the same y coordinate and

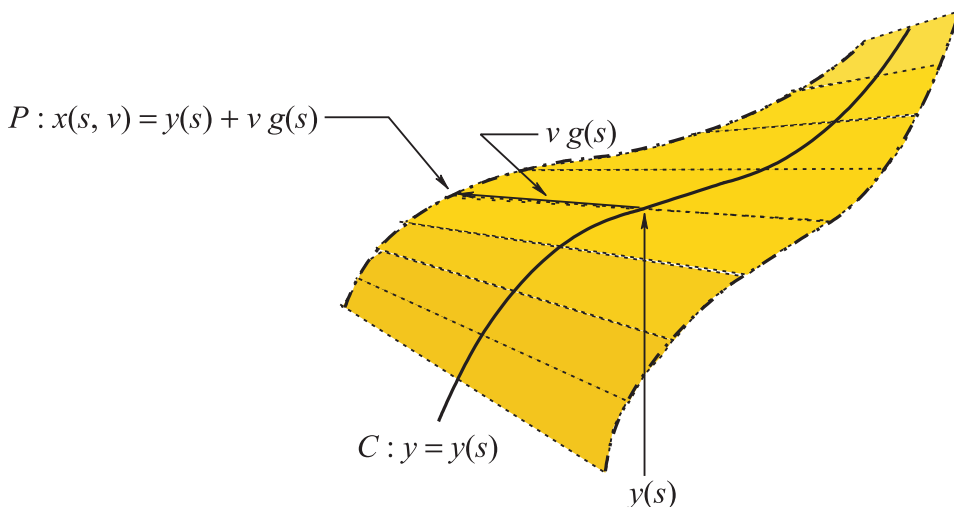


Figure 9.34. Generation of a ruled surface.

Notice the rulings successively “twisting” about the base curve C .

- (b) a bitangent plane is represented by crossing points. Find the number and positions of the crossing points.
18. By allowing “small variations” in the equation of developables (like the proximate planes in Chapter 8), define and study families of “approximate developables.” This is a serious *research problem*.

9.5 ♣ FT-9 Ruled Surfaces

Developable surfaces are a subset of the more general class of *ruled* surfaces \mathcal{R} that are generated by a one-parameter family of lines. A ruled surface, also called a *scroll*, can be created by the motion of a line called a *generating line*, *generator*, or *ruling*, as shown in Fig. 9.34. Let $C : \mathbf{y} = \mathbf{y}(s)$ be a curve and $\mathbf{g}(s)$ a unit vector in the direction of a ruling passing through a point $\mathbf{y}(s)$ of C . A point $P : \mathbf{x}(s, v)$ on the surface is then given by

$$\mathbf{x}(s, v) = \mathbf{y}(s) + v\mathbf{g}(s) \quad (\mathbf{y}' + v\mathbf{g}') \times \mathbf{g} \neq \mathbf{0} \quad \forall(s, v). \quad (9.73)$$

The curve $C : \mathbf{y} = \mathbf{y}(s)$ is called the *base curve* or *directrix*. The rulings are the curves $s = \text{constant}$. When \mathbf{y} is a constant, the surface is a cone, and when $\mathbf{g}(s)$ is a constant, the surface is a cylinder. The difference here is that $\mathbf{y}' + v\mathbf{g}' \times \mathbf{g} \neq \mathbf{0}$, allowing ruled surfaces to “twist,” unlike developables. The tangent plane π_P at a

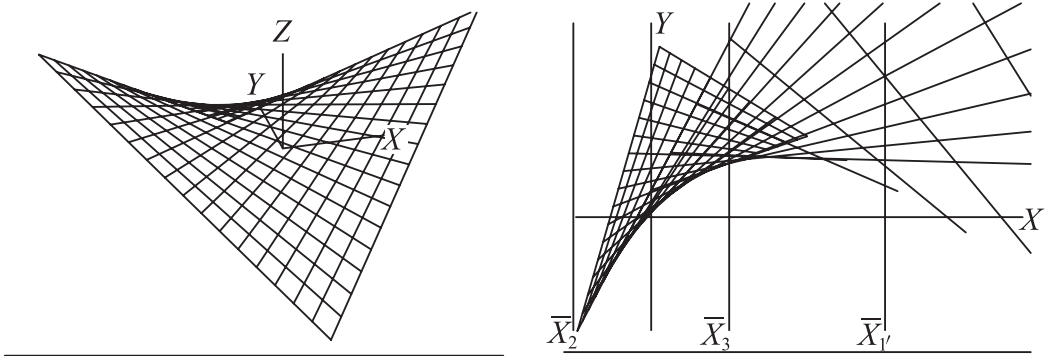


Figure 9.35. The saddle (left) is a doubly ruled surface. One of the two regions representing it on the left. Note the conic (parabolic) boundary.

point P on a ruled surface ρ contains the whole ruling r_P on P . For another point $Q \in \rho$, the tangent plane π_Q still contains r_P . Whereas all points on a ruling of a developable have the *same* tangent plane, moving along points on a ruling r of a ruled surface causes the corresponding tangent planes to *rotate* about r , this being the essential difference and the basis for the \parallel -coords representation of ruled surfaces.

There is a wonderful collection of ruled and many other kinds of surfaces in [130]. Ruled surfaces used as architectural elements are not only beautiful but also have great structural strength. A surface can be *doubly ruled* in the sense that any one of its points is on *two* lines completely contained in the surface (i.e., it can be generated by either one of two sets of moving lines). An elegant result [18, p. 42] is that in \mathbb{R}^3 , the *only* doubly ruled surfaces are quadrics: the hyperboloid of one sheet and saddle (hyperbolic hyperboloid).

It has been noted, in Exercise 15 in the previous section, that the representing curve $\bar{\sigma}_0$ of a developable σ can be obtained as the *envelope* of the family of lines \bar{R}_0 on the points $\bar{r}_{12}, \bar{r}_{13}, \bar{r}_{23}$ for each ruling r of σ . Similarly, $\bar{\sigma}_{1'}$ is the envelope of the lines $\bar{R}_{1'}$ on $\bar{r}_{1'2}, \bar{r}_{1'3}, \bar{r}_{23}$. A matched pair of points $\bar{\pi}_0 \in \bar{\sigma}_0, \bar{\pi}_{1'} \in \bar{\sigma}_{1'}$ represents the *single* plane containing $r \subset \sigma$ and tangent to every one of its points. By contrast, a plane tangent to a ruled surface ρ at a point contains a full ruling r , but as the point of tangency is translated continuously along r , the tangent plane continuously rotates about r . In \parallel -coords the points $\bar{\pi}_0, \bar{\pi}_{1'}$ continuously *translate* in tandem (with the same y coordinate) along the corresponding lines $\bar{R}_0, \bar{R}_{1'}$ representing the rotating tangent planes along r (see Fig. 5.6). An example is the *saddle*, Fig. 9.35 (left). By Corollary 9.3.5, it is represented by two regions whose boundaries are conic curves. The first region and the lines \bar{R}_0 tangent to the parabolic boundary (their envelope) are shown on the right.

Theorem 9.5.1 (Representation of Ruled Surfaces, A) *A ruled surface ρ is represented by the regions $\bar{\rho}_{j'}$, $j = 0, 1$, containing the families of lines $\mathcal{R}_j = \{\bar{R}_j\}$ whose envelopes are the boundaries $\partial\bar{\rho}_{j'}$ specific by Lemma 9.3.1. ♣ FT-9e*

To emphasize, a developable is represented by two curves that are the envelopes of the families of lines $\mathcal{R}_{j'}$ (formed from the representation of its rulings as described above). A ruled surface ρ is represented by two regions $\bar{\rho}_{j'}$, $j = 0, 1$, whose boundaries $\partial\bar{\rho}_{j'}$ are also the envelopes of the lines $\mathcal{R}_{j'}$ (obtained in the same way from the representation of its rulings), *together* with the lines $\mathcal{R}_{j'}$. Equivalently, it is helpful to consider $\partial\bar{\rho}_{j'}$ as a *line-curve* and $\bar{\rho}_{j'}$ the region covered by its tangents. The region's structure enables the construction of the matching algorithm for choosing pairs of points representing the surface's tangent planes. By the way, the boundary curves deriving from the rulings of a developable *differ* from those obtained from the rulings of a ruled surface. These points are clarified with the examples starting with the representation of the SA ,

$$SA : x_1^2 - x_2^2 = x_3. \quad (9.74)$$

We know that $\partial(\overline{SA})_{0'} = (SA \cap \pi_{0'}^s)_{12}$, and also that *one* projection of the space curve $SA \cap \pi_{0'}^s$ suffices. As indicated by the subscripts 12, here the projection on the x_1x_2 principal 2-plane is used. From $\pi_{0'}^s : x_1 - 2x_2 + x_3 = 0$, $x_3 = 2x_2 - x_1$ and upon substitution in (9.74) the desired curve is obtained:

$$(SA \cap \pi_{0'}^s)_{12} : x_1^2 - x_2^2 + x_1 - 2x_2 = 0,$$

a hyperbola whose representation in $\|\text{-coords}$ (see (9.20), (9.21), (9.22)) is as already mentioned, Fig. 9.35 (right), a parabola given by

$$(\overline{SA})_{0'} : -x^2 - 4y^2 - 4xy + 4x - 4y - 1 = 0. \quad (9.75)$$

The transformation of the hyperbola (9.5) to the parabola in (9.75) is nicely found, as shown in Fig. 9.36 (left), using a curve plotter.⁴⁶ The second region is found by repeating the procedure for $\pi_{1'}^s : x_1 + x_2 - 2x_3 = 0$, $x_2 = 2x_3 - x_1$, yielding

$$(SA \cap \pi_{1'}^s)_{1/3} : 2x_1^2 - 4x_3^2 + 4x_1x_3 - x_3 = 0.$$

This is a hyperbola, shown in Fig. 9.36 (right), whose transform in $\|\text{-coords}$ is the hyperbola

$$(\overline{SA})_{1'} : x^2 + 48y^2 - 16xy + 2x - 8y + 1 = 0, \quad (9.76)$$

⁴⁶Thanks to E. Rodity and V. Gurvich, who wrote the curve plotter for their course project

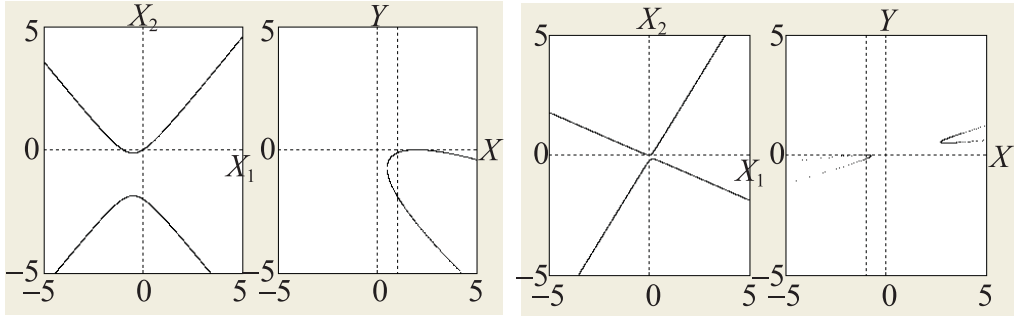


Figure 9.36. The intersection of the saddle SA and the first superplane π_0^s is a hyperbola.

(Left) The hyperbola transforms to a parabola, the boundary of the region $\overline{(SA)}_0$. (Right) The intersection of the saddle SA and the second superplane π_1^s is a hyperbola that transforms to another hyperbola, the boundary of the region $\overline{(SA)}_1$.

seen on the right. Here, as explained for (9.27), the (directed) interaxis distance $d = -1$ from the \bar{X}_1 to the \bar{X}_3 axis is used. These two regions are combined to provide the complete representation of SA . Prior to that, we rotate the x_1, x_2 axes about the x_3 axis by -45° , simplifying the saddle's equation to

$$SR : 2x_1x_2 = x_3 \quad (9.77)$$

and changing the planes tangent to the line u , which determine the presence of ideal points on the boundary curves, yielding the saddle's representing regions: $\partial(\overline{SR})_0$ is a hyperbola and $\partial(\overline{SR})_1$ an ellipse seen on the right of Fig. 9.38 after the -45° rotation.

To keep the subsequent discussion intuitive and reasonably precise without getting bogged down in technical details, we state without proof that the regions encountered have boundaries consisting of a finite number of simple (i.e., non-self-crossing) curves, each of which can be partitioned into a finite number of convex (upward or downward) curves. *Jordan's theorem* states that a simple closed curve separates the plane \mathbb{R}^2 into two regions: the interior and exterior. For our needs, we generalize this result to \mathbb{P}^2 by allowing a finite number of ideal points on the boundary which we consider “projectively” as a simple closed curve. Of course, the presence of ideal points needs to be treated judiciously. Relative to a region Ω , a point is *exterior* if there is a tangent from it to the boundary. Otherwise, it is *interior*. For $\partial\Omega$ simple and closed, there is a tangent in each direction, except those of the ideal points it contains. Hence, other than these, all other ideal points are exterior to Ω . Two points are on opposite sides of $\partial\Omega$, one interior and the other exterior, if the line on the points crosses the boundary an *odd* number of times. To avoid

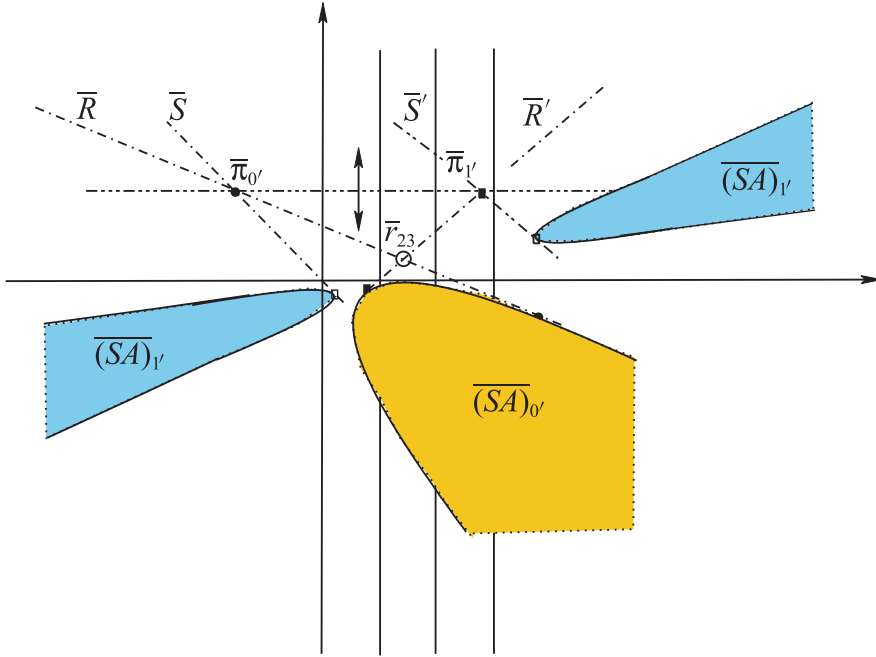


Figure 9.37. The saddle's representation is the complement of the shaded regions. The saddle SA is represented by the complements of the two shaded regions $(\overline{SA})_{0'}$ and $(\overline{SA})_{1'}$ having parabolic and hyperbolic boundaries respectively. The points $\bar{\pi}_{0'}$, $\bar{\pi}_{1'}$ representing a tangent plane and the ruling $r \subset \pi$ are constructed with the matching algorithm. A different ruling on π can be constructed suggesting that SA is doubly ruled.

complicating this rule, the line is not allowed to cross at points where two portions (one convex upward and the other downward) of $\partial\Omega$ are joined. For $\partial\Omega$ a convex curve (not closed), it is helpful to consider it as a line-curve to recognize that its exterior is still the region covered by its tangents. It is clear from Theorem 9.5.1 that for each point $\bar{\pi}_{j'} \in \partial\bar{\rho}_{j'}$, the region $\bar{\rho}_{j'}$ must contain at least a segment of the tangent at $\bar{\pi}_{j'}$, and hence all its points are on $\partial\bar{\rho}_{j'}$ or its exterior.

Corollary 9.5.2 (Representation of ruled surfaces; B) *A ruled surface $\rho \subset \mathbb{R}^3$ is represented by $\partial\bar{\rho}_{j'}$, $j = 0, 1$, and the regions exterior to it.*

Basically, this is a restatement of Theorem 9.5.1. Keep in mind that no points $\bar{\pi}_{j'}$ (representing tangent planes of ρ) can be in the *interior* of the regions whose boundaries are $\partial\bar{\rho}_{j'}$. Returning to SA , on the xy plane its representing regions $(\overline{SA})_{0'}$, $(\overline{SA})_{1'}$ consist of the boundaries and the *exterior* of the convex portions shown in Fig. 9.37 covered by their tangent lines. The two regions are *superimposed*

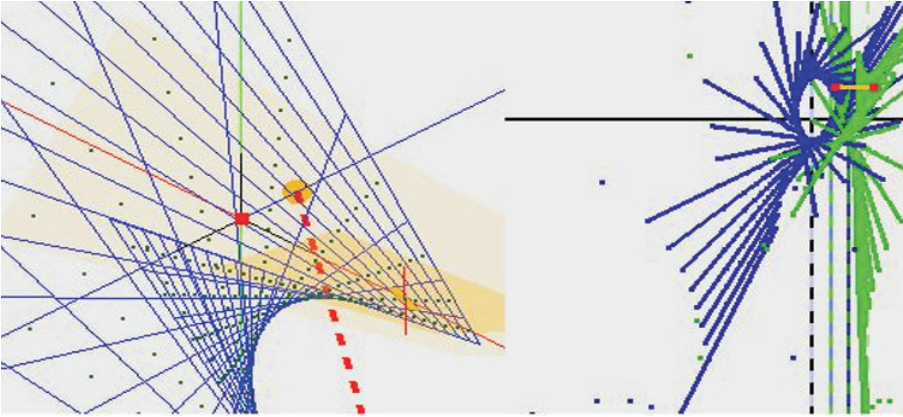


Figure 9.38. Conics representing the saddle.

Representation of the saddle showing an elliptical boundary and the tips of the two hyperbolic branches. For ruled surfaces, the regions are covered by the line tangents to the two boundaries. The two linked points represent a tangent plane.

in the sense that two indexed points, each belonging to a different region as distinguished from the index $0'$ or $1'$, may be in the same position (x, y) . As for developables, the boundaries $\partial(\overline{SA})_{0'}$, $\partial(\overline{SA})_{1'}$ are the envelopes of the family of lines derived from their rulings. Here the two boundary curves do not have the same y -range, something that distinguishes them from representing curves of developables. For the representation of a ruled surface ρ , the two regions $\bar{\rho}_{j'}$ need to be *linked* with a “matching algorithm” enabling the construction of valid pairs of points representing tangent planes. The algorithm is described in terms of the representation of SA in Fig. 9.37. On a line $y = c$, a point $\bar{\pi}_{0'} \in \bar{\rho}_{0'}$ is chosen and lines \bar{R}, \bar{S} from it tangent to each of the $\partial\bar{\rho}_{j'}$ are constructed. In general, there are several tangents from a point to each region, though here there are two tangents to each conic region. Choose any other point $\bar{\pi}_{1'} \in \bar{\rho}_{1'}$ on the line $y = c$, fixing the value of the first coefficient in π 's equation, also with two tangents \bar{R}', \bar{S}' to $\partial\bar{\rho}_{j'}$ on it. Then π , represented by $\bar{\pi}_{0'}, \bar{\pi}_{1'}$, is a tangent plane of SA , since the tangent lines generate the regions' boundaries (envelope). A ruling r on π is found by $\bar{r}_{23} = \bar{R} \cap \bar{R}'$ and another by $\bar{s}_{23} = \bar{S} \cap \bar{S}'$. This is significant, for as previously mentioned, SA is doubly ruled. However, there are several such intersections, suggesting that for developables, a convention is needed for the unique reconstruction such as the possible generalization of the *start points* convention for developables. Given $\partial\bar{\rho}_{j'}$, a line $y = c$ and a *link-function* $g(x_0, x_1)$ is provided, ensuring that for all values of x_j where the points are $\bar{\pi}_{j'} = (x_j, c)$ are exterior to

or on $\partial\bar{\rho}_{j'}$, the pairs $\bar{\pi}_{0'}, \bar{\pi}_{1'}$ represent tangent planes of ρ . Referring to Fig. 9.37, an additional rule $h(x_0, x_1)$ is required to match the tangents such as \bar{R} on $\bar{\pi}_{0'}$ with a tangent \bar{R}' on $\bar{\pi}_{1'}$. Then reconstruction of ρ 's tangent planes proceeds from the pairs $\bar{\pi}_{0'} \in \bar{R}, \bar{\pi}_{1'} \in \bar{R}'$ found by the vertical translation of the line $y = \text{constant}$ within the range of allowable values of y . The choice of the *start line* $y = c$ is important for its intersections with the $\partial\bar{\rho}_{j'}$; the points $\bar{\pi}_{j'}$ must be exterior to or on boundaries. Filling out the necessary details, providing a proof or refutation with counterexamples or alternative and better conditions, is left for Exercises 10 and 9.

The vertical translation of the line $y = \text{constant}$, while the points $\bar{\pi}_{j'}, j = 0, 1$, are constrained to slide along their tangents \bar{R}, \bar{R}' , corresponds to the rotation of the tangent plane about the ruling r . The tangent to $\partial\bar{\rho}_{j'}$ from $\bar{\pi}_{j'}$ represents a point $\pi_{j'}^s = \rho \cap \pi$ for *either* $j = 0$ or 1 so it is a point of tangency of π with ρ . The construction for finding the point of tangency of π with ρ when the $\bar{\pi}_{j'}$ are elsewhere on \bar{R} or \bar{R}' is deferred to the reader (Exercise 11). Avoiding tedious technicalities at the expense of some rigor, we outlined above the steps required to obtain the representation of a ruled surface ρ , the regions $\bar{\rho}_{j'}$, and the matching algorithm linking them. The “straight-lines pattern” within the regions $\bar{\rho}_{j'}$, characteristic of ruled surfaces, is easily discovered and explored interactively, as shown in Fig. 9.38, by rotating and translating the surface and viewing its resulting representation. A few words about the computation are in order. A rectangular grid is constructed in the x_1x_2 plane (or whatever parameters occur in the surface's description) and the tangent planes are computed at the grid points. The grid point density is also interactively adjusted to reduce clutter.

Conoids are ruled surfaces whose rulings are parallel to a plane and are perpendicular to an *axis* normal to the plane such as those shown in Fig. 9.39.⁴⁷ Spiral staircases are shaped after the *right conoids*. Another specimen is the *parabolic conoid* seen in Fig. 9.40 (center), which, like the saddle, has a parabolic directrix. It is given by

$$PC : x_1x_3^2 = x_2, \quad (9.78)$$

and is defined in the first and third quadrant of the x_1x_2 plane. The boundaries of its representing regions are the images of

$$\begin{aligned} \partial(\overline{PC})_{0'} &= (PC \cap \pi_{0'}^s)_{12} : x_1^3 - 4x_1^2x_2 + 4x_1x_2^2 - x_2 = 0, \\ \partial(\overline{PC})_{1'} &= (PC \cap \pi_{1'}^s)_{1'3} : x_1x_3^2 - 2x_3 + x_1 = 0. \end{aligned} \quad (9.79)$$

These are cubic curves, and from Plücker's formula, Section 7.6, we find that their transforms are algebraic curves of degree 6. The first is a cardioid (left) in Fig. 9.40,

⁴⁷The figures were obtained using the beautiful software written by M. Spivak for her course project.

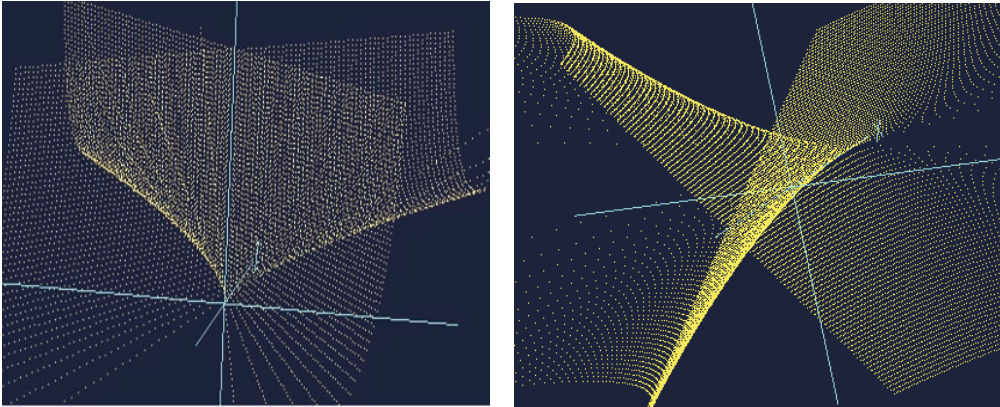


Figure 9.39. Conoids.

The rulings of Whitney's umbrella (left) and Plücker's conoid (right) are perpendicular to a line, which here is coincident with the vertical axis. They are both *conoids*.

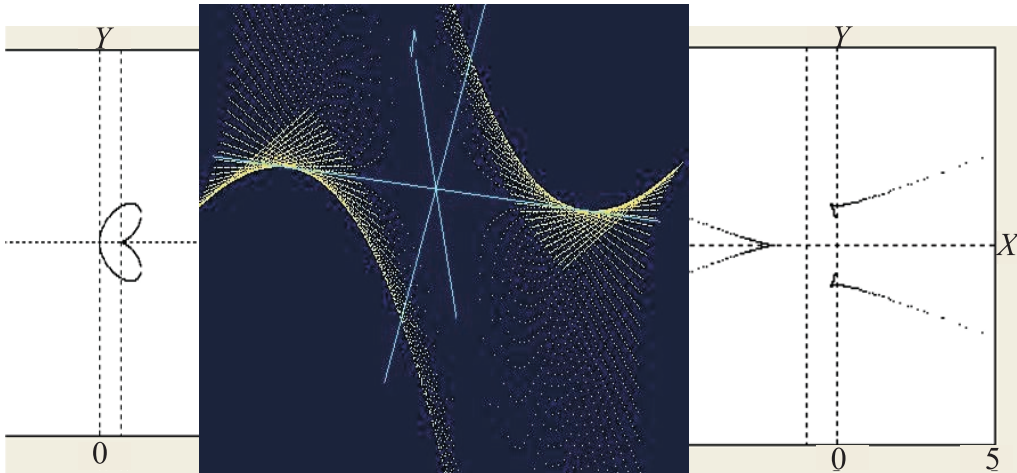


Figure 9.40. (Center) The parabolic conoid PC is a ruled algebraic surface of degree 3.

The first representing region $(\overline{PC})_{0'}$, a cardioid, is on the left, and the second $(\overline{PC})_{1'}$ on the right. The cusps are the images of inflection points.

and the second is on the right. Both are symmetric with respect to the x axis, the cusps being the transforms of inflection points. The region $(\overline{PC})_{0'}$ consists of the cardioid's exterior, the region covered by its tangents, and $(\overline{PC})_{1'}$ is the exterior of the six convex curves forming its boundary. The superposition of the two representing

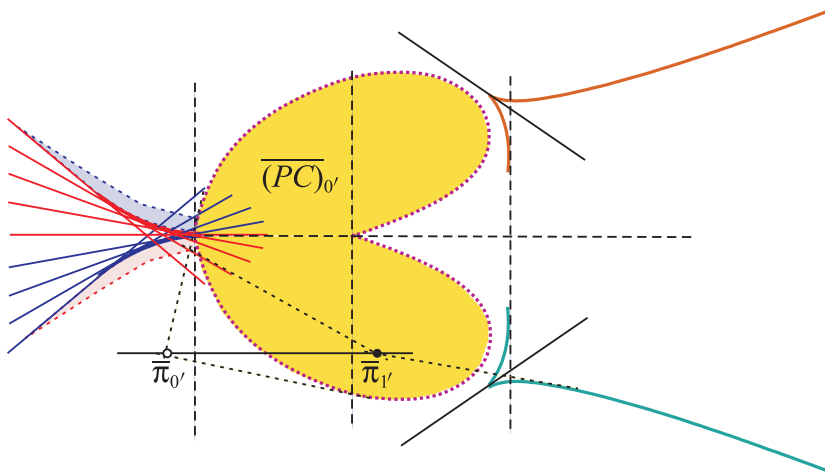


Figure 9.41. Representation of the parabolic conoid and one of its tangent planes.

regions is shown in Fig. 9.41, where some tangents on the left cusp are drawn illustrating subregions where the exterior of some boundary curves overlap. The regions are linked with the matching algorithm described above. On a line $y = c$, a point $\bar{\pi}_{0'} \in \overline{(PC)}_{0'}$ is chosen having two tangents to the cardioid's boundary. The second point $\bar{\pi}_{1'} \in \overline{(PC)}_{1'}$, also on $y = c$, may have several tangents to its region's boundaries, reminding us of the need, mentioned earlier, for a convention enabling the unique reconstruction of a ruled surface from its representation (Exercise 10). The representation picture is like a landscape map pointing out interesting features and pathways inviting exploration and study of the surface. The goal is, as we gain experience and understanding, the discovery of invariants corresponding to distinct surface properties such as dualities for the developables. In Fig. 9.42 are the representations for two orientations with the straight lines tangent to the regions' boundaries.

A famous ruled surface is the *Möbius strip*, described by

$$\mathbf{x} = \mathbf{y}(\theta) + v\mathbf{g}(\theta), \quad -\frac{1}{2} < v < \frac{1}{2}, \quad 0 \leq \theta \leq 2\pi,$$

$$\mathbf{y}(\theta) = (\cos \theta)\hat{\mathbf{e}}_1 + (\sin \theta)\hat{\mathbf{e}}_2, \quad (9.80)$$

$$\mathbf{g}(\theta) = \left(\sin \frac{1}{2}\theta \cos \theta \right) \hat{\mathbf{e}}_1 + \left(\sin \frac{1}{2}\theta \sin \theta \right) \hat{\mathbf{e}}_2 + \left(\cos \frac{1}{2}\theta \right) \hat{\mathbf{e}}_3,$$

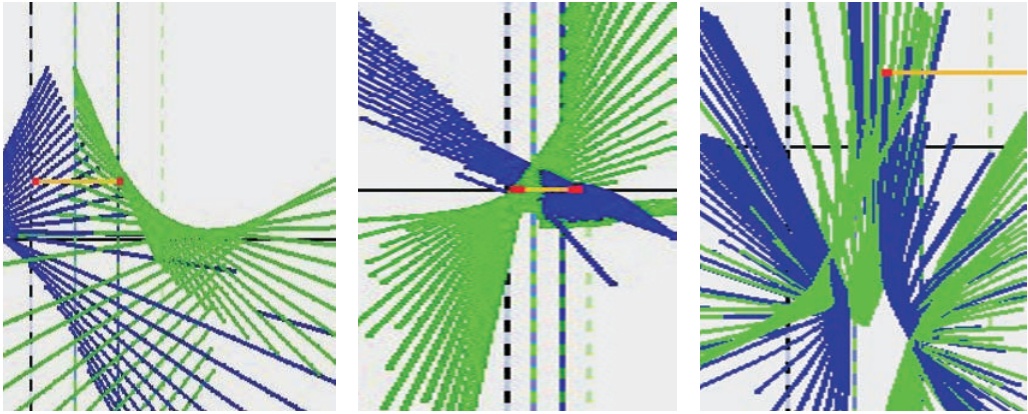


Figure 9.42. Representations (left and center) of a *parabolic conoid* for two orientations. Representation of the *ruled cubic* (Exercise 4) on the right. The straight lines reveal that the surfaces are ruled and partially outline the regions' boundaries. A tangent plane is represented by the two linked points.

It is *nonorientable* in the sense that tracing the normal vector at a moving point along a loop, specified by its directrix the circle $\mathbf{y}(\theta) = \cos \theta \hat{\mathbf{e}}_1 + \sin \theta \hat{\mathbf{e}}_2$ [128, p. 170], the normal flips by 180° from its original direction when the traversal is completed. In other words, this surface has only one side. The strip's structure is elucidated in Fig. 9.43 (right), showing that the ruling moving along the directrix twists at an angle $\theta/2$ from the vertical, where θ is the angle swept along the directrix. The ruling intersects the central axis in three positions, inverting its orientation by 180° by the time it completes the circuit to create the twisted wonder we see in Fig. 9.43 (left).

How does this surface appear in $\|\text{-coords}$? Dazzling patterns represent it at various orientations, Fig. 9.44. Of particular interest is the search for a pattern characterizing nonorientability. In Fig. 9.45 we see a closed circuit on the strip and the corresponding pair of representing curves. This is repeated for a different orientation, producing a pair of representing curves, that are easier to study, Fig. 9.46 (right). Each intersection point represents a tangent plane π for which $\tilde{\pi}_0'$ and $\tilde{\pi}_1'$ coincide, that is, π is perpendicular to the x_2x_3 principal plane with its first coefficient $c_1 = 0$. A nice way to understand the twisting is suggested in Fig. 9.46 (left), where the strip is placed on a plane, in this case the x_2x_3 plane, creating three “folds” that do not have to be symmetrically positioned like those in the picture. At each of the folds, the tangent plane is perpendicular to the x_2x_3 plane, where by the “flip” corollary, Corollary 5.3.2, the first coefficient c_1 of the tangent plane's equation changes sign, say $+\rightarrow -\rightarrow +\rightarrow -$, from the start to the completion of the circuit.

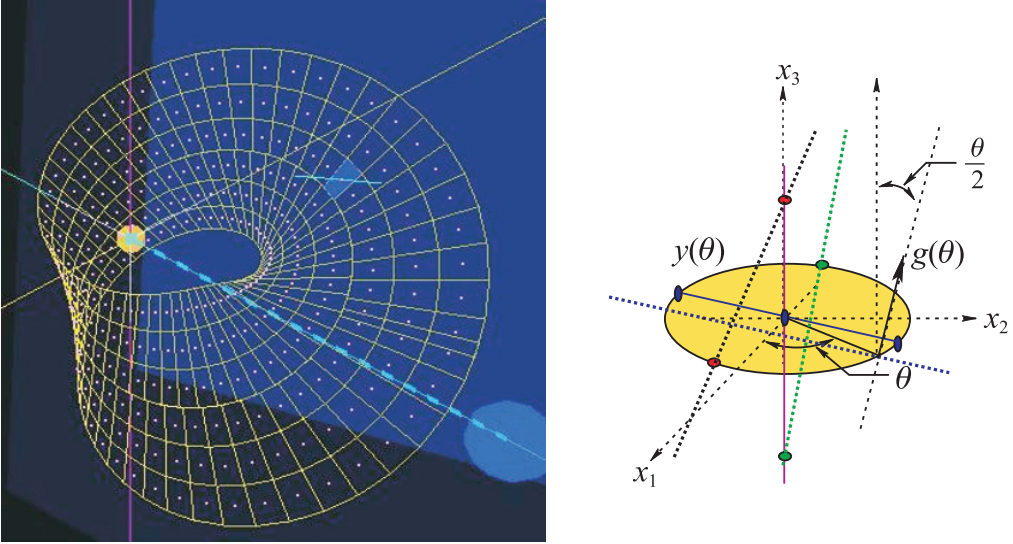


Figure 9.43. *Möbius strip* surface and its structure (left).

A ruling traversing the circular directrix intersects the central axis three times in one complete circuit.

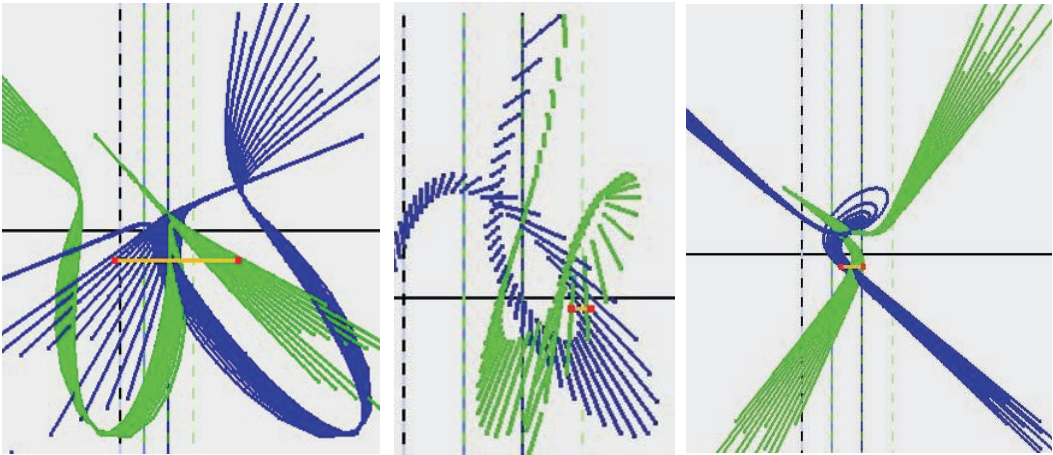


Figure 9.44. The two cusps on the left (dual of inflection points, see Fig. 9.23) show that the Möbius strip has an “inflection point” in three dimensions. The curves going towards infinity upwards and downwards show that it is a closed surface. The cusps and loops may merge as shown on the right. A tangent plane is represented by the two linked points.

At each of the three intersections of the representing curves in Fig. 9.46 (right), $c_1 = 0$. By the same argument, the remaining coefficients c_2, c_3 vanish three times

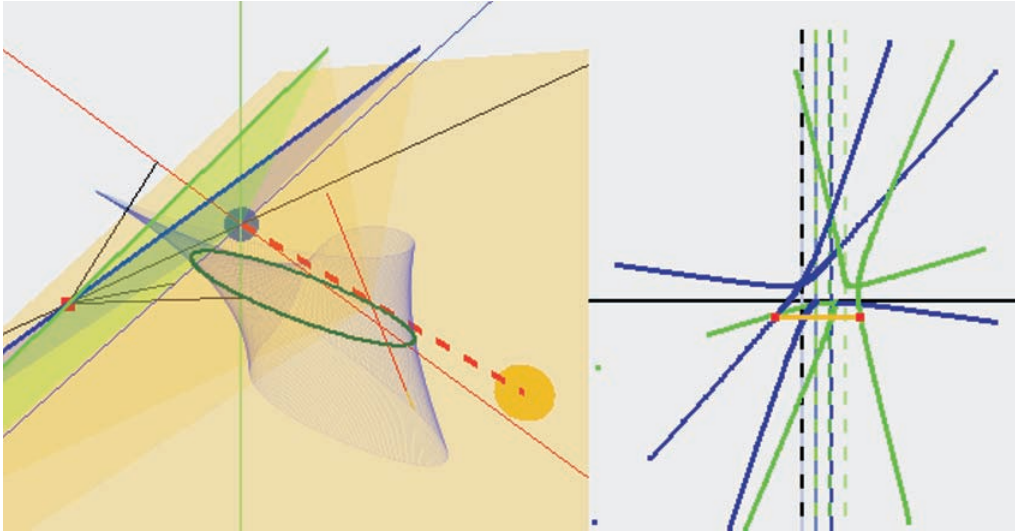


Figure 9.45. Traversing one circuit on a Möbius strip.
Note the intersections between the pair of representing curves.

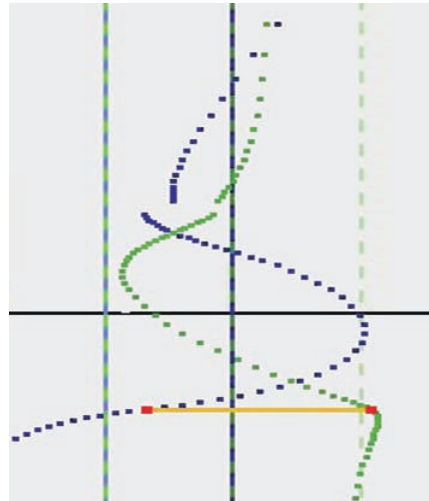
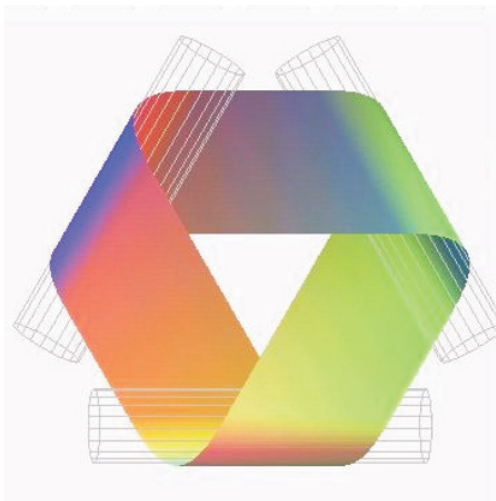


Figure 9.46. Visualizing nonorientability.

(Left) *Möbius strip* (thanks and acknowledgment to [130]) placed on one of the principal planes, say x_2x_3 . Note the three folds resulting in three tangent planes perpendicular to the x_2x_3 plane. (Right) Traversal of a closed circuit on the *Möbius strip* represented in \parallel -coords by two curves. There are three crossing points, each corresponding to a tangent plane perpendicular to the x_2x_3 plane.

and change their sign. Hence the tangent plane's normal vector $\vec{c} = (c_1, c_2, c_3)$ becomes $-\vec{c}$ and flips direction. If the triple intersection of the representing curves is characteristic of nonorientability, what about the more numerous intersections in the pair of curves in Fig. 9.45? It is in such situations that interactivity is essential. Choosing a start position with a pair of linked points representing a tangent plane, the curves are traversed, where we observe that *the two points do not coincide at every intersection point but only at three of them*. With a bit of patience, the reader can actually verify this from the picture, starting the traversal from the indicated position with the two linked points and tracing their path, paying attention to the three ideal points in each representing curve, returning to the start, observing that $\bar{\pi}_0'$ and $\bar{\pi}_1'$ coincide at only three of the intersection points. The three intersections together with the ensuing flip are clearly a necessary and sufficient condition for nonorientability. For a closed circuit on a doubly twisted Möbius strip, the representing curves $\bar{\pi}_0'$ and $\bar{\pi}_1'$ coincide at six intersection points corresponding to six folds, and so on.

It appears, then, that a pattern characteristic of nonorientability has been found whose verification relies on interactively tracing a closed circuit on the representing curves. The picture, of course, is not proof, but it provides insights that may lead to a proof; this is an objective of visualization. This points the way to researching nonorientability in \mathbb{R}^4 and higher dimensions as well as to knot theory. Here is another thought: twisting a curve about its tangent at a point creates an inflection point. So is the cusp seen in Fig. 9.44 (right), which is the dual of an inflection point, together with the surrounding pattern in the second curve, an inkling of an nonorientable twist, like a “3-D inflection point” produced by twisting and then joining a curved strip? Though imprecise, this depiction is in the spirit of visualization, being intuitive and stimulating ideas. Does the birdlike pattern represent the Möbius strip's nonorientability? These observations and questions may stimulate further research on the representation of nonorientability and ruled surfaces. The representation results thus far can be summarized in the following theorem.

Theorem 9.5.3 (Representation of developable and ruled surfaces in \mathbb{R}^N) *A surface $\sigma \subset \mathbb{R}^N$ is represented by $N - 1$ regions $\bar{\sigma} \subset \mathbb{P}^2$ and is*

1. **ruled but not developable** \Leftrightarrow *all points of $\bar{\sigma}$ are exterior with respect to the boundary of at least one of its $N - 1$ regions,*
2. **developable** \Leftrightarrow *$\bar{\sigma}$ consists of $N - 1$ curves, the boundaries of $\bar{\sigma}$, all having the same y-range.*

Prior to treating more general surfaces, an older inexact but very useful surface representation is included, illustrated with an application to instrumentation and

process control. It is of specialized interest and independent, so that it may be skipped without impairing other chapter topics.

Exercises

1. Prove, or give a counterexample, that a ruled surface that is *not developable* cannot be *convex*.
2. Prove theorem 9.5.3 for \mathbb{R}^N .
3. Construct the \parallel -coords representation of the hyperboloid of one sheet, Fig. 9.3, given by $x_1^2 + x_2^2 - x_3^2 = 1$.
4. Construct the representation of the ruled cubic $x_3 = x_1^2 x_2$.
5. Construct the representation of *Zindler's conoid* given by $2x_1 x_2 = x_3(x_1^2 - x_2^2)$.
6. Construct the representation of *Plücker's conoid* given by $x_3(x_1^2 + x_2^2) = x_1^2 - x_2^2$; Fig. 9.39 (right).
7. Construct the representation of *Whitney's umbrella* given by $x_3^2 = x_2^2$; Fig. 9.39 (left).
8. Explore the representation of the *parabolic conoid* in Fig. 9.41. List and justify your findings.
9. Determine from its representation $\bar{\rho}$ whether the surface ρ is doubly ruled.
10. Provide a convention, like the “start points” for developables, that enables the unique reconstruction of a ruled surface ρ from its representation; this is a *research problem*.
11. Provide an algorithm for constructing the point of tangency of a plane π tangent to a ruled surface ρ from the representations $\bar{\rho}$ and $\bar{\pi}$; this is a *research problem*.
12. Prove or provide a counterexample of the nonorientability pattern conjecture for \mathbb{R}^3 and \mathbb{R}^N ; this is a *research problem*.

9.6 ** Approximate Quadric Hypersurfaces

The problem of finding interior points to hypersurfaces is important in applications such as process control and decision support [81]. Here an algorithm for finding and *visualizing* such points is given. In the interest of brevity, the algorithm is discussed concurrently with the representation of some of the hypersurfaces to which it applies. The point is constructed by successively defining its coordinates in some order. The following steps are required:

1. Initially find the *range* of the variable whose value is going to be selected.
2. Select a value of the variable in this range. This reduces the dimensionality of the point selection problem.

3. Find the range of the next variable for the reduced convex hypersurface and select a value in that range.
4. Repeat the previous step until the last variable is selected.

The convexity and boundedness of the convex hypersurface guarantee that in each case, the range is an interval. The success of the algorithm hinges on finding the intervals efficiently.

As we will see, for bounded convex hypersurfaces the intervals are found by taking the projections of the hypersurface's sections. The boundary of these sections is transformed into parallel coordinates, resulting in the stepwise display of the point selection process. Such a display provides:

1. information on the *local* properties of the hypersurface and
2. *proximity* information for the selected point, i.e., a measure of distance of the point from the boundary.

These points are illustrated subsequently when the algorithm is applied to some specific kinds of hypersurfaces. The algorithm also applies to certain nonconvex hypersurfaces where the available ranges are a *finit union of intervals* at every stage of the point selection process.

9.6.1 The Hyperellipsoids

Let A be a positive definite symmetric matrix and consider the function

$$F(X) = (X, 1) \begin{pmatrix} A & e^T \\ e & \alpha \end{pmatrix} \begin{pmatrix} X^T \\ 1 \end{pmatrix}, \quad (9.81)$$

where $X = (X_1, X_2, \dots, X_N)$ and e are row vectors, and α is a constant. The equation

$$F(X) = 0$$

defines a hyperellipsoid with real points X , provided that

$$e(A^{-1})^T - \alpha > 0.$$

A point is interior, on, or exterior to the surface according as $F(X)$ is less than, equal to, or greater than 0. Given a point X , it is straightforward to determine its location relative to the hyperellipsoid. The object here is to find, component by component, an interior point.

There are two kinds of processes required for this purpose. Fixing the value of one of the variables produces a planar section of the hyperellipsoid that is

itself a hyperellipsoid of dimension one less than the original; this necessitates the computation of a new array for the resulting reduced hyperellipsoid. The other requirement is to calculate the available range of the remaining free variables by taking the projection of the reduced hyperellipsoid on a coordinate plane. Now let

$$A = \begin{pmatrix} A_1 & B^T \\ B & C \end{pmatrix}, \quad e = (e_1, e_2), \quad (9.82)$$

where A_1 consists of the first 2×2 submatrix of A , C is A with the first two rows and columns deleted, B and B^T account for the remainder of A , with e_1 consisting of the first two elements of e and e_2 , which is the remainder of e . It is a relatively straightforward to prove the following.

Theorem (B. Dimsdale). *Let $X^* = (X_1, X_2)$. The projection of $F(X^*)$ on the X_1, X_2 plane is the ellipse*

$$f(X^*) = 0 = (X^*, 1) \begin{pmatrix} \bar{A} & \bar{e}^T \\ \bar{e} & \bar{a} \end{pmatrix} \begin{pmatrix} X^{*T} \\ 1 \end{pmatrix} \quad (9.83)$$

and

$$\left. \begin{aligned} \bar{A} &= A - B^T C^{-1} B, \\ \bar{e} &= e_1 - e_2 C^{-1} B, \\ \bar{\alpha} &= \alpha - e_2 C^{-1} e_2^T. \end{aligned} \right\} \quad (9.84)$$

Given the projection, the maximum ranges of X_1 and X_2 are easily calculated and so is the available range for X_2 , given X_1 .

The hyperbola that is the transform of the ellipse into parallel coordinates is essential for the display of the selection process. For one thing, the available range of X_1 is the interval on the X_1 axis between the two branches of the hyperbola, and similarly for X_2 . Furthermore, the *range* of X_2 , once X_1 is selected, is *exhibited* by drawing tangents to the hyperbola from the point on the X_1 axis, the chosen value of X_1 . As the process is iterated through all the variables, the entire sequence of selections and ranges is presented in a parallel coordinate display. Furthermore, the hyperbolas themselves provide indications as to the sensitivity of various choices. For an example of the display, see Fig. 9.47. In that picture the outer hyperbolic segments are the transforms of the original hyperellipsoid projections on the coordinate planes. The inner segments come from the successive slices of the hyperellipsoid and show the successive ranges. Note that a point is exterior to the hyperellipsoid if and only if the intermediate envelopes, representing successive slices of the hypersurface, are crossed by the polygons, is interior if it nowhere touches, and is on the surface otherwise. The computational complexity is $O(N^3)$ [42].

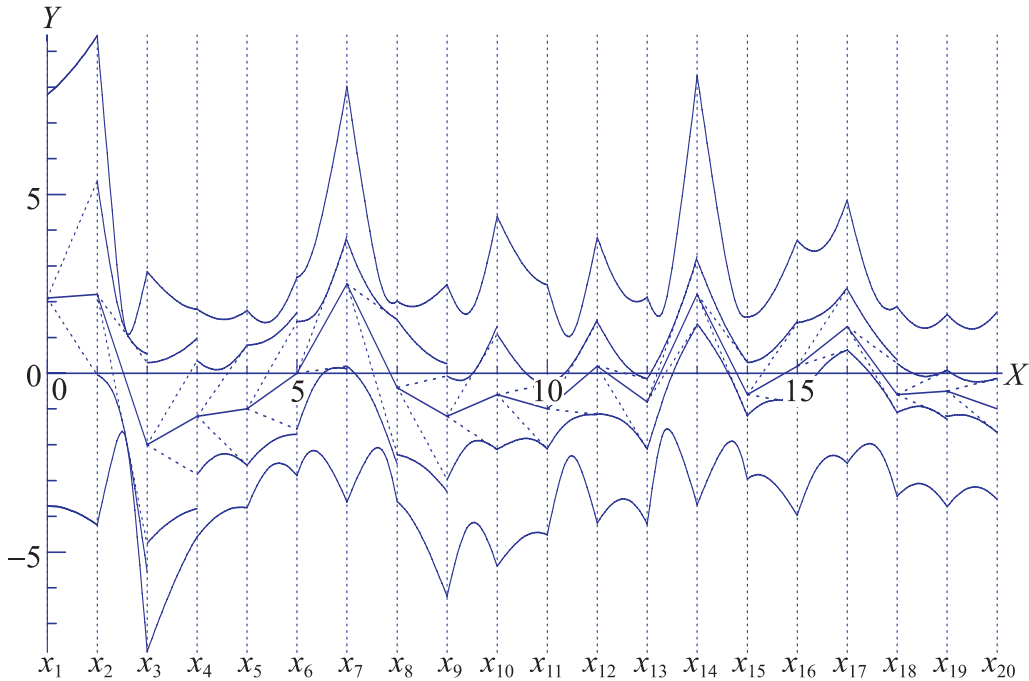


Figure 9.47. Interior point construction for a convex 20-dimensional hyperellipsoid.

This can be the model of a process depending on 20 parameters. An interior point is constructed and represented by the polygonal line shown. The values of the 20 parameters specify a *feasible* state of the system. At this state the *critical parameters* are x_1^3, x_1^4, x_1^5 where the available ranges (seen between the two inner curves) are narrowest, i.e., the point is closest to the boundary.

9.6.2 The Hyperellipse

The procedure for hyperellipsoids calls for repeated projections on coordinate planes, at a considerable cost in computational complexity. The obvious suggestion is to define an object whose projections do not change as the iteration proceeds.

Consider a collection of hypercylinders whose projections on the coordinate planes are ellipses. The intersection of such a set of hypercylinders, if it exists, is clearly a convex object. With some attention to detail, such objects can be defined so that they exist, are bounded, and have a property called *consistency*, which, briefly, means that a valid value for *any* variable is consistent with a valid value for *any other* variable. A notable property of this kind of surface is that the computational complexity is $O(N)$, since the surface is defined by $N - 1$ ellipses, whence $N - 1$ quadratic equations must be solved [42].

9.6.3 Intelligent Process Control

Representation of the Domain

The following points are worth noting at the outset of this section:

1. Controlling a process requires considerable human intelligence. A proposed system, or any other, is called *intelligent* if it can replace—to a considerable extent—the human operator for this task, or if an operator is present, the system can render *intelligent help* by choosing the *critical* parameters, at any given time, with explanations as to why they are critical. Further, there is the ability to do *what if scenarios* and see what options there exist for getting out of “tight spots.”
2. It is known that any *dynamical* system with N parameters can be considered as a *static* system in $N + 1$ parameters; the original N parameters *and time*. This applies also to processes.

An approximate but useful representation of a general convex hypersurface can be obtained by computing the *envelope* of the collection of polygonal lines representing its points. For multidimensional polyhedra, such a representation is *unique*. Otherwise, the description of the hypersurface needs to be retained in addition to its parallel coordinate representation. Such a representation of a convex hypersurface in \mathbb{R}^{20} is shown in Fig. 9.47. It corresponds to a particular (nonlinear) relation, which for our purpose describes a *process* involving 20 parameters labeled x_1, \dots, x_{20} . An N -tuple is a *feasible point for the process* iff it is *interior to the hypersurface*. This is the *geometric equivalent* of the statement that a specific combination of parameter values satisfies all the constraints imposed on the process. When a mathematical description for the process is not available, an approximation (a best fit according to some criterion) for the envelope can be obtained from available data. Geometrically, the relation between adjacent axes is indicated by the shape of the envelope, and here, due to the hypersurface’s convexity, is a *gh* (a generalized hyperbola). To economize on-display space, only the strip between the axes is shown. Further, the envelope’s shape clearly depends on the particular order in which the variables appear (as axis labels). As we will see, rather than a problem, this provides an opportunity for choosing *preferred orderings of the variables* subject to certain criteria.

The Rule

Controlling the process is equivalent geometrically to staying in the interior of the hypersurface. We apply the algorithm for doing so. The interval on the x_1 axis between the upper and lower portions of the envelope indicates the range of parameter x_1 . When a value for $x_1 = a_1$ is chosen in the available range of x_1 , the

number of variables in the relation is reduced by one (i.e., from 20 to 19 variables in our example). Again with reference to Fig. 9.47, from the fixed value of x_1 , tangents to the upper and lower portions of the envelope are drawn (here the points of tangency are not seen, since they lie outside the strip between the x_1 and x_2 axes). The envelope of the 19-dimensional hypersurface is then obtained from the (description) of the original hypersurface with the value $x_1 = a_1$. Again for display-space economy, only the strip between x_2 and x_3 of the *new* envelope is shown and is always contained between the upper and lower portions of the original envelope. The *restricted available range* of x_2 , due to the constraint of fixing the value of one of the variables, is the interval on the x_2 axis between the upper and lower tangents. Proceeding, a value $x_2 = a_2$ is chosen within the *current available range* of x_2 , reducing the dimensionality by one (e.g., 18), and the whole process is repeated. A polygonal line is shown that always lies between the intermediary envelopes. It represents an interior point to the hypersurface, and *all* interior points can be found in this way. So this provides us with a way of finding and *displaying feasible points for the process*. Also the algorithm is very fast and suitable for real-time operation.

The Display as an Instrument Panel

When the value of a variable is fixed, the *complete* envelope of the 19-dimensional surface may be drawn, showing at a glance not only the current value of any of the parameters (in this case x_1) but, unlike a conventional “instrument panel,” the *current available range of all the parameters*; this is due to the interrelationship among them. There are other salient features of the display. Note that for the parameters x_{13} , x_{14} , x_{15} the available ranges are the narrowest. This indicates that the feasible point is *closest to the boundary* with respect to these *critical* parameters. Hence, when there is need for rapid decisions (such as in piloting very fast aircraft), significant information reduction can be achieved by controlling the critical parameters whose *priority is decided by their proximity to the boundary*, adjusting the remaining variables subsequently. This also brings out the possibility of displaying or highlighting only the information desired (possibly specific subsystems) at any given time. When a point is not feasible, the display shows the corresponding polygonal line crossing some of the intermediate envelopes. It is then evident *which parameters must be changed and by how much* in order to bring the polygonal line inside the envelope and the process in control. In short, this instrument panel provides a variety of information that an operator would somehow need to know (at least implicitly derive or guess based on experience) in order to control the process.

Order, Learning, and Explanations

Ideally, one would like to have a display in which the relation between any pair of variables, say x_i and x_j , is apparent. There are $N(N - 1)/2$ such pairs for N variables, which suggests that $O(N^2)$ displays with different orderings may be needed. Fortunately, this is not the case. It turns out [184] that for $N = 2n$, exactly n “well chosen” displays suffice, while for $N = 2n + 1$, $n + 1$ displays suffice. This is due to the fact that a complete graph of N vertices can be obtained as a union of that many “properly chosen” spanning trees. For example, for $N = 6$ the permutations 126354, 231465, 342516 circularly contain every possible pair (independent of the order) of *adjacent subscripts from 1 to 6*. Concerning the ordering of the variables, an advantageous ordering is according to sensitivity. In fact, it is possible to implement a dynamic ordering in which the order at each step is decided according to some rule, for example, that the variable displayed and fixed next is the one with the smallest available range.

When an approximation of the hypersurface is obtained from data, the representation can be altered in view of new information. When a new feasible point is found that is not obtained via the algorithm from the approximate representation of the hypersurface, using the approximation technique again would enlarge the hypersurface in some region. On the other hand, an incorrect feasible point obtained from the approximation would yield a restriction of the hypersurface. So a continuously more accurate representation of the process can be obtained as a result of successive trials. Hence, the system has the ability to *learn*.

At any stage of the algorithm an *explanation of why* a specific value of a variable results in a feasible or infeasible point is available in terms of whether the polygonal line intersects the intermediate envelopes. Also, having arrived at a partial choice of parameter values, it is possible to conduct *what if scenarios*—an important look-ahead feature—before actually fixing the value of the remaining parameters. What is proposed then is a new technique for process control, which also suggests a *generalization of the concept of expert system* to include those whose *knowledge representation* is a hypersurface together with a rule for “navigating” in the interior.

9.7 **FT-10** More General Surfaces

The formulation in (9.15) applies to any sufficiently smooth surface. We open this last section by examining the representation of the venerable sphere in \mathbb{R}^3 ,

$$SP : x_1^2 + x_2^2 + x_3^2 = r^2. \quad (9.85)$$

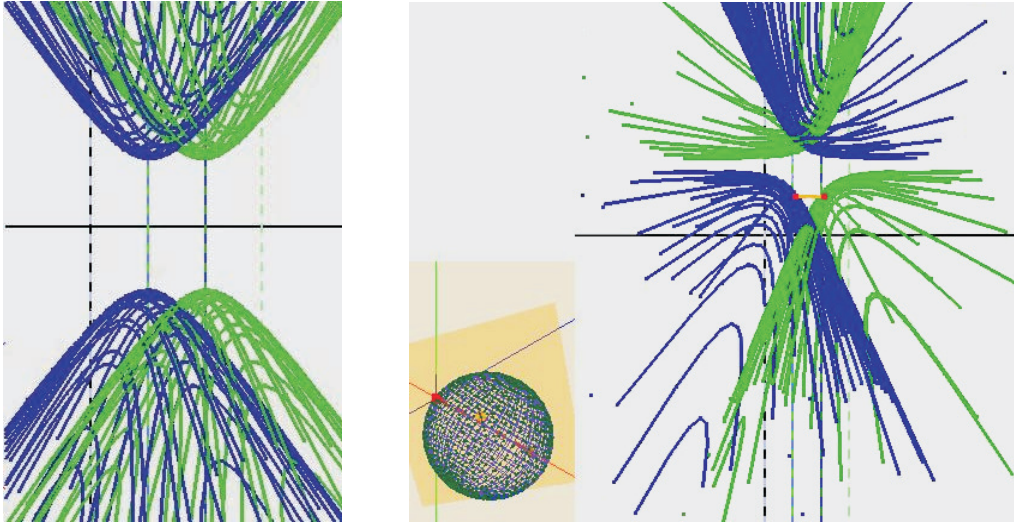


Figure 9.48. Representation of a sphere.

Representation of a sphere centered at the origin (left) and after a translation along the x_1 axis (right), causing the two hyperboloids to rotate in opposite directions.

Its intersections with the first superplane π_0^s , a circle $C0$, and its hyperbolic image $\overline{C0}$, are

$$C0 : 2x_1^2 - 4x_1x_2 + 5x_1^2 - r^2 = 0,$$

$$\overline{C0} = \partial(\overline{SP})_{0'} : 3r^2x^2 - 6y^2 - 6r^2x + 5r^2 = 0. \quad (9.86)$$

The hyperbola's vertices (i.e., min and max) are at the points $(1, \pm\sqrt{2/3}r)$. The second boundary $\partial(\overline{SP})_{1'}$ is the same hyperbola translated to the right by one unit, as shown on the left of Fig. 9.48. Due to its symmetry, a hypersphere in \mathbb{R}^N , like a hypercube, Fig. 9.1, is represented by $N - 1$ identical patterns (hyperboloids) spaced successively one unit apart. Translation of the sphere along the x_1 axis rotates $\partial(\overline{SP})_{0'}$ clockwise and $\partial(\overline{SP})_{1'}$ counter-clockwise, Fig. 9.48 (right), an interesting manifestation of the *translation* \leftrightarrow *rotation* duality we have encountered on previous occasions. It is useful to compare this representation with the inexact one given in the previous section. Consider the 5-D sphere

$$5SP; x_1^2 + x_2^2 + x_1^3 + x_4^2 + x_5^2 = r^2, \quad (9.87)$$

whose projection on the principal x_1x_2 plane is the circle $x_1^2 + x_2^2 = r^2$ represented by the hyperbola $2r^2x^2 - y^2 - 2r^2x + r^2 = 0$ going through the points

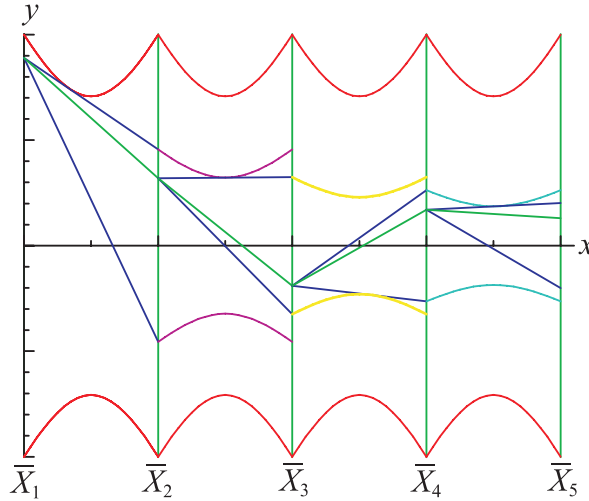


Figure 9.49. Interior point construction of a sphere in \mathbb{R}^5 .

$(x = 0, y = \pm 1)$ and $(x = 1, y = \pm 1)$. The other projections are represented by translations of this hyperbola successively by one unit. All four are superimposed as in Fig. 9.49, showing that the exact representation, also with hyperbolic boundaries, is an “upgrade” though the hyperbolas differ. The picture helps in following the construction of a point interior to $5SP$. Let us start by choosing a value of $x_1 = a_1 \in [-r, r]$, placing this on the \bar{X}_1 axis, and drawing the tangents from this point $(0, a_1)$ to the upper and lower branches of the first hyperbola. This is equivalent to intersecting $5SP$ with the plane $x_1 = a_1$, resulting in the 4-D sphere

$$4SP : x_2^2 + x_1^3 + x_4^2 + x_5^2 = r^2 - a_1^2, \quad (9.88)$$

with radius $\sqrt{r^2 - a_1^2}$, seen in the figure on the \bar{X}_2 axis at its intersection with each of the two tangents. Also shown is the hyperbola representing the projection of $4SP$ on the x_2x_3 plane. A choice $x_2 = a_2 \in [-\sqrt{r^2 - a_1^2}, \sqrt{r^2 - a_1^2}]$ is made, reducing the dimension, providing a 3-D sphere, and so on. As long as the successive choices are in the allowable range (i.e., radius of reduced sphere), the polygonal line joining the a_i does not intersect the intermediate curves like the one shown, and it represents an interior point of $5SP$. If the polygonal line is tangent to any of the intermediate curves, it represents a point on the surface, and if it intersects, then it represents an exterior point. It is clear, then, that all interior and surface points of a sphere $SPN \subset \mathbb{R}^N$ are represented by the polygonal lines between the hyperbolas that represent the projections on the principal planes. Therefore, these hyperbolas

slightly overlap the hyperbolas $\partial(\overline{SP}_i)$ that intersect the \bar{X}_i axes at $y = \pm\sqrt{2/3}r$ (Exercise 4).

All the points representing tangent planes are in the *interior* of the upper and lower branches (Exercise 5). Recall the unit vector $\hat{\mathbf{u}} = 1/\sqrt{3}(\hat{\mathbf{e}}_1 + \hat{\mathbf{e}}_2 + \hat{\mathbf{e}}_3)$ along the line u . Let π_u be the plane through the origin with unit normal $\hat{\mathbf{u}}$. The intersection $\pi_u \cap SP = CU$ is a great circle, which plays an important role in the representation of SP . All tangent planes of SP along CU have the sum of their coefficients $S = c_1 + c_2 + c_3 = 0$ vanishing, for $\mathbf{N} \cdot \hat{\mathbf{u}} = 0$, where $\mathbf{N} = (c_1, c_2, c_3)$, since being parallel to the line u , they are represented by ideal points, thus causing the split into two hyperbolic branches. The representing points of the tangent planes above CU are on the branch, and those below form the lower branch. Note that at the north and south poles the tangent planes are $\pi_p : x_3 = \pm 1$ with representing points $\bar{\pi}_{0'} = (2, \pm r, 1)$ on the \bar{X}_2 axis but *not on the boundary* $\overline{SP}_{0'}$, which is formed by the representing points of tangent planes on $\pi_{0'}^s \cap SP$ and does not include the poles. Similarly, the $\bar{\pi}_{1'}$ are on the $\bar{X}_{1'}$ axis and the same $y = \pm r$.

Whereas the interior points of the hyperbolic regions properly matched (Exercise 6) represent the sphere's tangent planes, there is another interesting matching that gives a feel about the sphere's interior. Consider the representation of concentric spheres (centered at the origin) with radii r ranging between a minimum r_m and maximum r_M . As r increases, the hyperbolas' vertices and the rest of the curve rise linearly with r , enlarging the region between the upper and lower branches. Let a pair of points at $y = y_a$ be given on the hyperbolic boundaries representing a tangent plane π_a of one of the spheres. Its radius a is found from the vertices of the hyperbolas with extreme values at $y = \pm\sqrt{2/3}a$. Draw the vertical lines on the pair of given points. Any other pair of points at $y = y_c$ represents a tangent plane π_c parallel to π_a on the sphere with radius $c = y_c/y_a a$.

The intersection of the superplanes with ellipsoids and paraboloids are ellipses, whence their representation by hyperbolic regions, Fig. 9.50. For the paraboloid, the intersections may also be parabolas, in which case the representing regions would be either hyperbolas, each with one vertical asymptote, or vertical parabolas; see Section 7 (Exercise 7). In general, the intersection of the superplanes with bounded and convex surface σ is a bounded convex set bc whose $\|\cdot\|$ -coords representation is a generalized hyperbola gh , and such an example is shown on the left in Fig. 9.51 (Exercise 8).

With the stipulation that all representing points are in the *interior* or boundary points, consider a bounded convex hypersurface whose representation at a specific orientation consists only of gh regions. By the *translation* \leftrightarrow *rotation* duality evident in Fig. 9.48, the regions remain gh for all other orientations (i.e., translations and/or rotations). The representation of an unbounded convex hypersurface also consists

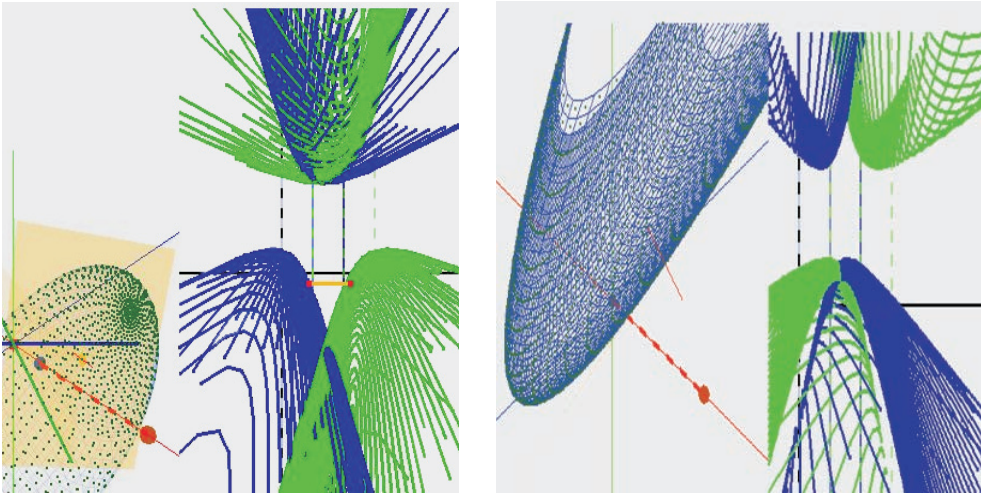


Figure 9.50. Representation of an ellipsoid on the left and a paraboloid on the right.

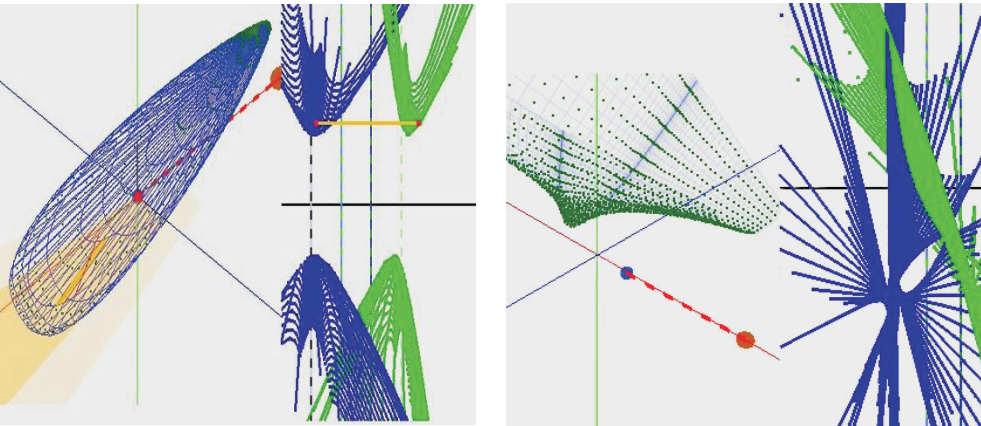


Figure 9.51. Representations: non-quadric convex surface (left) and a surface with a small bump (right).

of gh regions unless it has some regions (say p) in the direction of u , in which case p of its representing regions are parabola-like gp (generalized parabolas)⁴⁸ having one ideal point and $N - p - 1$ gh regions. Without entering into details, we conjecture that the gp regions have ideal points in the vertical direction.

⁴⁸Discussed in the chapter on curves.

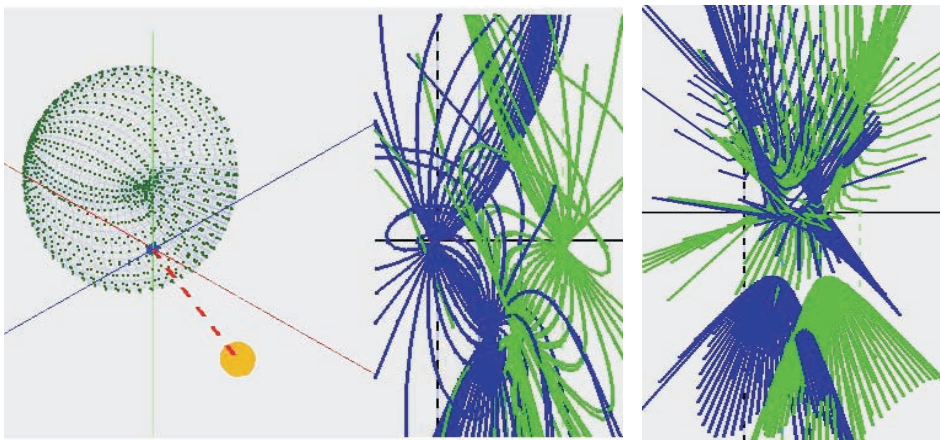


Figure 9.52. Representation of a nonconvex surface with two pointed depressions.

The depressions are mapped into pairs of “swirls.” Each planar section of the depression is a cusp and by the *cusp – inflection-point* duality (see Fig. 9.23 and 9.24) is mapped into two curves with inflection-points, which together form the pair of swirling patterns. Both depressions are visible in the representation though one is hidden in the perspective on the left.

Theorem 9.7.1 (Convexity) *A bounded surface $\sigma \subset \mathbb{R}^N$ is convex \Leftrightarrow for one orientation it is represented by $N - 1$ gh regions whose points are interior or on the boundary. If the representation has some gp and the rest gh regions, then the hypersurface is unbounded and convex.*

This is a powerful result, important in many applications, enabling *recognition of convexity in \mathbb{R}^N* . Next we examine the representation of surfaces having small deviations from convexity. In Fig. 9.51 on the right is a surface σ_b with a small “bump.” Its orientation was interactively varied in every imaginable way, and significantly, its representations, like the one shown, were all markedly different from a *gh*. In certain orientations the intersections of σ_b with the superplanes are *bc*, bounded and convex, yet the representation regions were not *gh*. This important observation is explored further. No less significant is that the representing regions are not chaotic, having interesting similar shapes varying with changing orientation (Exercise 9a). Proceeding, a surface σ_d with two small “dimples” or depressions, considered here as a perturbation of convexity “opposite” to that in σ_b , shown in Fig. 9.52, is explored next. Interesting representation patterns are seen with varying orientations; one (on the right) contains two hyperbolic regions and other patterns *outside* them, evidence of “near convexity.” The two pairs of like “swirls” (center) correspond to the two dimples (Exercise 9b). Again it is rather remarkable that for a multitude of orientations, none of the representations consists of *just* two hyperbolas, as may naively be expected when the intersection with the superplanes is a bounded convex set *bc*.

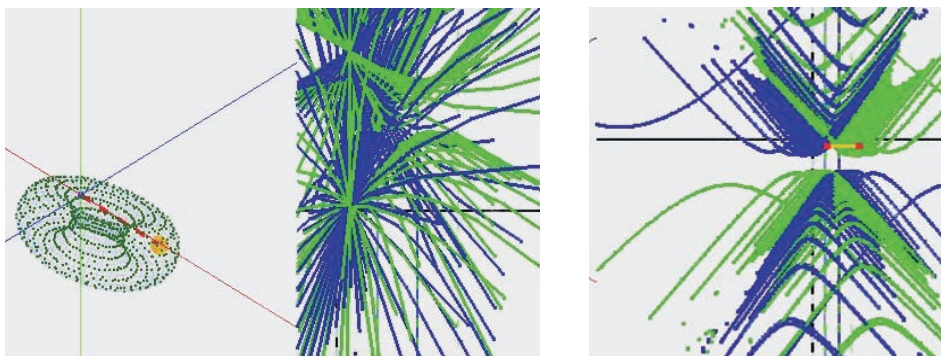


Figure 9.53. Torus representation showing two bitangent planes (left) and hyperbolic slices (right).

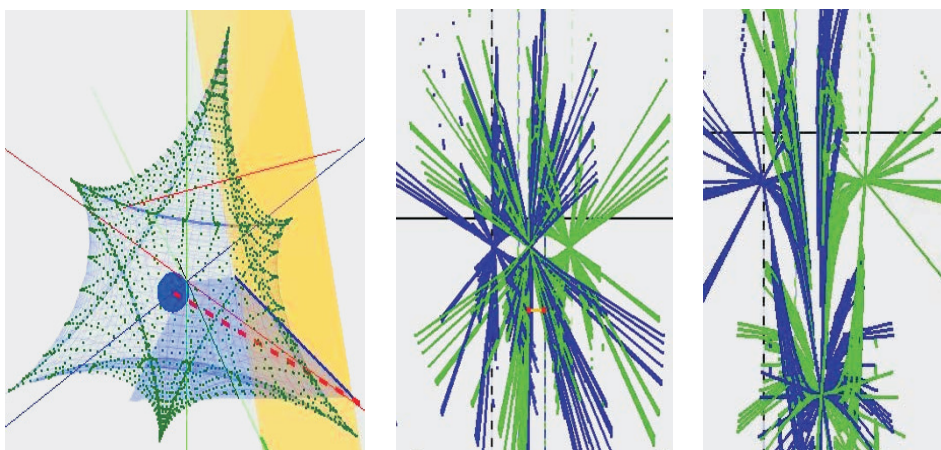


Figure 9.54. Astroidal surface and its representations for two orientations.

There are two pairs of crossing points (one pair is overlapping in the middle) which are present for all orientations.

Many hyperbolas appear in the representation of the torus, Fig. 9.53 (right), but unlike the representing hyperbolas of convex surfaces, here the hyperbolas are not contained within an outermost hyperbola but are spread in the horizontal direction. Curves from the two representing regions intersect at two points with the same x coordinate. This is reminiscent of the crossing points corresponding to bitangent planes. We surmise that the crossing points represent the two *parallel* planes tangent at the top and bottom of the torus (Exercise 10). The representation of the astroidal surface [130] in Fig. 9.54 has a pattern with three crossing points that persists in several orientations. Upon careful scrutiny we see that only at the midpoint do

curves from the two regions intersect, resembling the crossing points in the torus's representation. Could this point represent a tangent plane approaching one of the surface's vertices and becoming a bitangent plane in the limit? Another association that comes to mind is of an exterior supporting plane touching the surface at exactly three vertices.

The significant finding from these examples is that the $\|\cdot\|$ -coords representation even of complex nonconvex surfaces in various orientations is not chaotic but consists of discernible patterns associated with their properties. All this may lead to new surface classifications (Exercise 11). The surface representation seems preferable to the standard surface representation (in terms of projections) even for \mathbb{R}^3 ; surface properties like convexity and features like dimples, bumps, coiling and more are not hidden in various orientations. The reason is that in $\|\cdot\|$ -coords representation, a surface is not only described by its points but *also* by the **normal vector** at each point (review Fig. 9.2).

Arising from this discussion is the question, what happened to the claim that the representing boundaries are the images of the surface's intersection with the superplanes? Let us return to the proof of Lemma 9.3.1 and note that the argument is based on the assumption that the curve is on *one side of its tangent*. Recall that in Section 6.2, on envelopes, the analysis was also based on the family *being on one side of the envelope*, leading to necessary and sufficient conditions for the existence and a method for obtaining the envelope. As was seen in the subsequent examples, the method works even when there are singularities and when this condition fails. In such cases the envelope has several components, each consisting of special points on the family of curves, e.g., inflection points, cusps, extrema, crossing points, and others. For the problem now at hand, the boundary of a representing region is an envelope of lines "touching" it (but not necessarily tangent to it) and the curves formed by the representing points. We saw that the method of Lemma 9.3.1 applies well to developable, convex, and even *ruled nonconvex* surfaces; the representing regions lie on the interior or exterior respectively and *on one side of the boundary*. We found that for nonconvex surfaces, parts of the representing regions lie on both sides of the boundary, so the condition of Lemma 9.3.1 fails, and the boundary is not necessarily only the image of the surfaces' intersection with the superplanes. As pointed out below, this has important ramifications and is basically good news. The refinement of Lemma 9.3.1 for nonconvex surfaces is a difficult research problem, related to that for the envelopes of general families of curves and surfaces. These observations, even if partially true, may lead to a new approach for the construction of (also approximate) convex hulls in \mathbb{R}^N as for \mathbb{R}^2 [110].

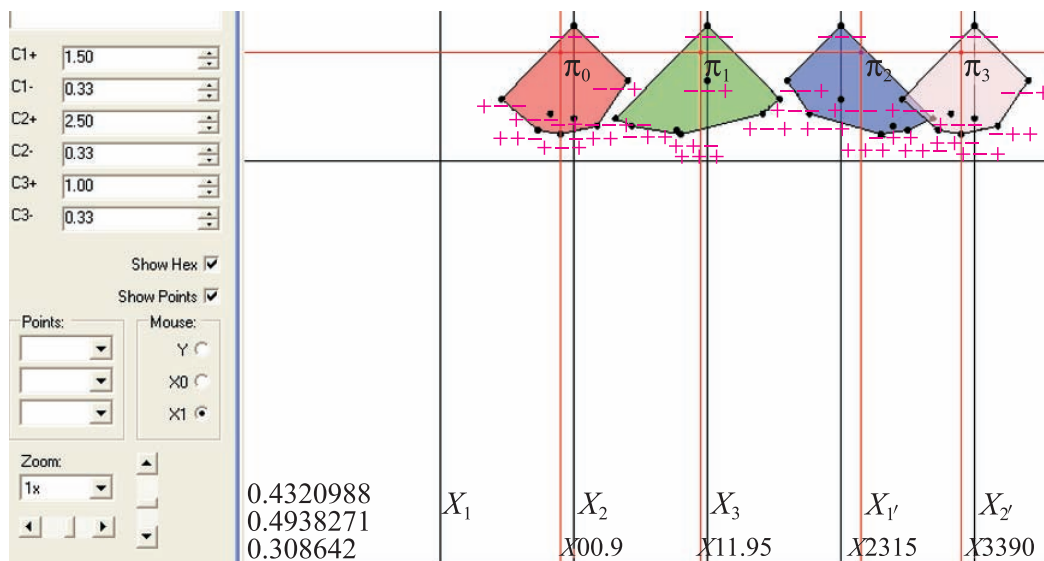


Figure 9.55. Hexagons representing a family Π of proximate planes.

The coefficients' range is $c_1 \in [1.5, 0.33]$, $c_2 \in [2.5, 0.33]$, $c_3 \in [1, 0.33]$, $c_0 = 1$ in \mathbb{R}^3 . The two points interior to the first two hexagons represent the plane with (see lower left) $c_1 = 0.43$, $c_2 = 0.49$, $c_3 = 0.31$, so it is not in the family Π even though the third point is interior to the third hexagon. The software used allows for the interactive variation of the coefficients and positions of the representing points in the first two hexagons.

9.7.1 ** A More General Matching Algorithm

Recall the family of curves gh (generalized hyperbola), uc (unbounded convex), and bc (bounded convex) in Section 7.7. For our purposes here, consider their generalization as the class CC of surfaces in \mathbb{R}^N such that for a $\sigma \in CC$, in a given orientation, the boundaries of the $N - 1$ regions of $\bar{\sigma}$ are either gh , uc , or bc , with the points representing their tangent planes interior to their respective regions. The matching algorithm scheme in Section 8.3.8, on proximate flats, can be used also for this class of surfaces. For this reason we recall the procedure, featuring another student course project.⁴⁹ As Fig. 9.55 shows, it does not suffice that $\bar{\pi}_0$ and $\bar{\pi}_1$ be in the interior (or boundary) of the first and second hexagons respectively for the plane $\pi \in \Pi$. The hexagons are bc bounded convex sets, and so the collection of planes in Π cannot be convex; for then it would be represented by gh regions.

⁴⁹Acknowledgment and thanks to Y. Lifshits and I. Makarov, who wrote this wonderful program for their course project in 2005: navigation within the hexagonal regions representing families of proximate planes.

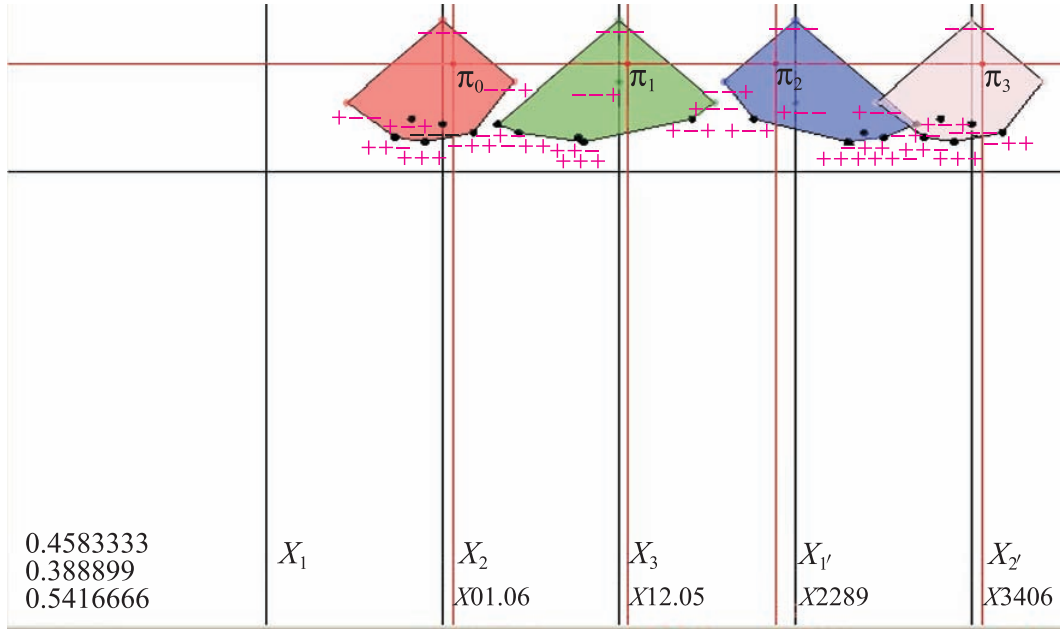


Figure 9.56. A plane in the family Π .

Here the two indicated points represent the plane with $c_1 = 0.458$, $c_2 = 0.389$, $c_3 = 0.542$, which is in the family Π . Three vertices are chosen representing a simplex in \mathbb{R}^3 and represented by triangles within the hexagons. Two representing points within the triangles in the first and second hexagons represent a plane in Π .

A simplex within Π is chosen by specifying three vertices in the hexagons. For $\bar{\pi}_0$ and $\bar{\pi}_{1'}$ in the triangles within the first and second hexagons, the pair represent a valid plane $\pi \in \Pi$, as shown in Fig. 9.56. What is helpful here is that a triangle in \mathbb{R}^2 is also represented by a triangle in $\|\cdot\|$ -coords. So the triangular faces of the 3-D simplex remain triangular in their representation. The process can be continued for higher dimensions,⁵⁰ allowing also variations in c_0 ; Fig. So $\Pi \subset \mathbb{R}^3$ is represented by octagons with two vertical edges, Fig. 9.57 and $\Pi \subset \mathbb{R}^4$ by decagons with two vertical edges.

In view of what we have learned, we can surmise that the *intersections* of the hyperplanes in a family $\Pi \subset \mathbb{R}^N$ of proximate hyperplanes form a *gh* if it is represented by convex $2N$ -gons, a *uc* if it is represented by *uc*'s, and a polyhedron if it is represented by *gh*'s. Care is required to state this precisely (Exercise 3). The reason for bringing up this approach again is that it may apply to surfaces in

⁵⁰Acknowledgment and thanks to A. Feigenbaum, T. Kafri, and S. Kamhin for extending the Y. Lifshits and I. Makarov software to \mathbb{R}^4 .

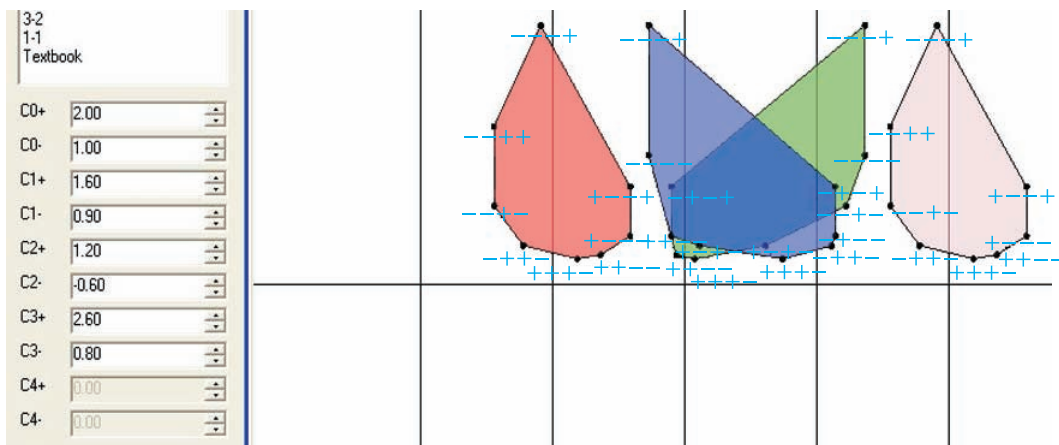


Figure 9.57. Octagons representing a family of proximate planes $\Pi \subset \mathbb{R}^3$. The two vertical edges are due to the variation of c_0 ; here $c_0 = [1, 2]$.

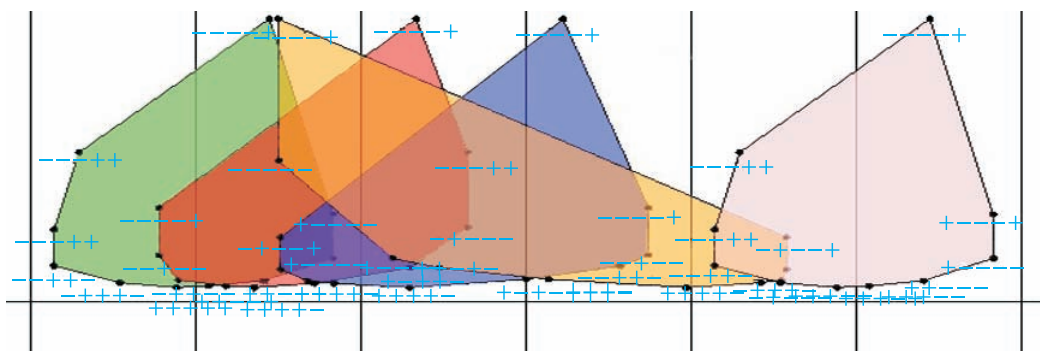


Figure 9.58. Decagons with two vertical edges representing proximate hyperplanes in \mathbb{R}^4 .

CC , whose representation consists of gh , uc , or bc regions with all the representing points in their interior, *and* the regions can be partitioned into convex portions (Exercise 12).

Tantalizing as it has been, we wind up for now our glimpse into the challenges, beauty, and promise of surface representation.

Chapter acknowledgments. While writing this chapter I recognized even more the crucial contributions of my colleagues and students. In the early 1980s, *Bernie Dimsdale*, my long-time friend and colleague, and I started studying the representation of surfaces summarized in Section 9.6. *Chao-Kuei Hung*, in the early

1990s, pioneered the exact representation of surfaces in terms of their tangent planes, obtaining results on developables and initiating the study of ruled surfaces. Many insights and beautiful figures were obtained with the wonderful interactive software written by *David Adjashvili* for his course project. Years earlier, *Michael Spivak* was capable and brave enough to write the first interactive software on surface representation. Liat Cohen's clever notation for the axis spacing together with many other ideas and suggestions were very helpful. There are others already acknowledged and those who contributed their software in the attached CD. To all I am profoundly grateful.

Exercises

1. (a) Prove that the hyperbolic boundaries of the sphere SP rotate in opposite angular directions with a translation of SP along the x_1 axis. Find the exact relation between the translation by Δx_1 and angle of rotation.
 (b) What changes in the representation with translations along the remaining axes?
 (c) Generalize for the representation a sphere in \mathbb{R}^N translated along an axis $x_i, i = 1, \dots, N$.
2. With reference to Fig. 9.1, find the representing boundaries of the cube in \mathbb{R}^3 using Lemma 9.3.1 and show that they are *gh* or *uc*. Generalize this for the hypercube in \mathbb{R}^N .
3. For a family of proximate hyperplanes $\Pi \subset \mathbb{R}^N$ describe precisely (with proofs) the boundary of Π when it is represented by $N - 1$ convex $2N$ -gons, *ub*, or *bc* regions.
4. Provide an algorithm for constructing interior points of a sphere directly from the hyperbolas of its exact representation; *hard*.
5. Prove that the points representing tangent planes of a sphere are each on the boundary or *interior* to its corresponding hyperbolic region.
6. Provide the matching algorithm for the sphere's representation; this is a *research problem*.
7. Under what conditions is an elliptic paraboloid represented by vertical parabolic regions or by hyperbolas each with one vertical asymptote?
8. Provide criteria for distinguishing
 - (a) the convex quadrics (cylinder, ellipsoid, paraboloid) from each other by their hyperbolic representing regions. Hint: the section on conics in Chapter 7 on curves;
 - (b) the orientation of the convex quadrics from their representation.

9. (a) Take a convex body such as a sphere in \mathbb{R}^3 , introduce a small bump on its surface, and derive the representing shapes of the resulting surface σ'_b ; *hard*.
(b) Take a convex body such as a sphere in \mathbb{R}^3 , introduce a small dimple, and derive the representing shapes of the resulting surface σ'_d ; *hard*.
(c) Compare the representations of σ'_b and σ'_d . Provide a criterion for distinguishing between the two kinds of nonconvexity.
10. Prove or disprove that the two tangent planes at the top and bottom of the torus are represented by two points at the same x coordinate where curves from the two representing regions cross; see Fig. 9.53, center.
11. Study the representation of the class of surfaces of revolution. Provide criteria for recognizing a surface of revolution from its representation; this is a *research problem*.
12. Apply the method in the last section to the representation of a sphere in \mathbb{R}^3 .
13. Prove or give a counter example of the *dimple* \rightarrow *swirl* duality seen in Fig. 9.52.

This page intentionally left blank



10.1 Introduction

The first, and still most popular, application of parallel coordinates is in exploratory data analysis (EDA) to discover data subsets (relations) that fulfill certain objectives and guide the formulation of hypotheses. A data set with M items has 2^M subsets, any one of which may be the one we really want. With a good data display, our fantastic pattern-recognition ability cannot only cut great swaths in our search through this combinatorial explosion, but also extract insights from the visual patterns. These are the core reasons for data visualization. With parallel coordinates (\parallel -coords), the search for relations in multivariate data sets is transformed into a 2-D pattern-recognition problem. Guidelines and strategies for knowledge discovery are illustrated on several real data sets, one with hundreds of variables. A geometric classification algorithm is presented and applied to complex datasets. It has low computational complexity, providing the classification rule explicitly and *visually*. The minimal set of variables required to state the rule is found and ordered by their predictive value. Multivariate relations can be modeled as hypersurfaces and used for decision support. A model of a country's economy reveals sensitivities, impact of constraints, trade-offs, and sectors unknowingly competing for the same resources. This chapter can be read independently of the rest, with foundational background provided where needed. Readers are encouraged to browse through the book, looking into specific topics when the motivation arises. Collision-avoidance algorithms for air traffic control are discussed separately in Chapter 4, on multidimensional lines.

10.1.1 Origins

For the visualization of multivariate problems, numerous mappings encoding multidimensional information visually into 2-D or 3-D (see [63] and [172], [173], [174]) have been invented to augment our perception, which is limited by our 3-dimensional habitation. Wonderful successes such as Minard's "Napoleon's March to Moscow," Snow's "dot map," and others are ad hoc and exceptional. Succinct multivariate relations are rarely apparent from *static* displays; *interactivity* is essential. In turn, this raises the issues of effective *GUI* (graphic user interface), queries, exploration strategies, and information-preserving displays, but we are getting ahead of ourselves.

10.1.2 The Case for Visualization

Searching a data-set with M items for interesting, depending on the objectives, properties is inherently hard. There are 2^M possible subsets, any one of which may satisfy the objectives. The *visual cues* that our eyes can pick from a good data display help us to navigate through this combinatorial explosion. How this is done is part of the story. Clearly, if the transformation $data \rightarrow picture$ clobbers information, a great deal is lost right at the start. We postulate that a display of data sets with N variables suitable for *exploration* satisfies the following requirements:

1. *should preserve information* – the dataset can be completely reconstructed from the picture,
2. *has low representational complexity* – the computational cost of constructing the display is low,
3. *works for any N* – not limited by the dimension,
4. *treats every variable uniformly*,
5. *has invariance under projective transformations* – the data set can be recognized after rotations, translations, scalings, and perspective transformations,
6. *reveals multivariate relations in the dataset* – the most important and controversial single criterion,
7. *is based on a rigorous mathematical and algorithmic methodology* – to eliminate ambiguity in the results.

No completeness or uniqueness is claimed for this list, which should invite criticism and changes. Commentary for each item, illustrated by examples and comparisons, is listed in the same order.

1. The numerical values for each N -tuple (i.e., each data point) are recoverable from the scatterplot matrix (SM) and the \parallel -coords display. This may not be necessary or desirable for presentation graphics (e.g., pie charts, histograms).

2. In the pairwise scatterplot matrix, the N variables appear $N(N - 1)/2$ times. By contrast, there are only N axes in \parallel -coords, though there is an additional preprocessing cost, pointed out in the ensuing. For $N \geq 10$, the practical barriers due to the required display *space* and visual difficulty limit the use of *SM*. These barriers are less stringent for \parallel -coords.
3. Even with a nice perspective, orthogonal coordinates are inherently limited to $N = 3$, due to the dimensionality of our existence. With added aspects, the illusion of a few more dimensions can be added. By contrast, for \parallel -coords, implementation capabilities rather than conceptual barriers determine the maximum useable N .
4. An example is the “Chernoff Faces” display, where each variable corresponds to a specific facial feature and treated accordingly. The correspondence *facial feature* \rightarrow *variable* is arbitrary. Choosing a different correspondence gives a different display, and the fun part is that there is no way to know that the *two different* displays portray the *same data set*. Of course, this is true for general *glyphs* displays.
5. This is true for *SM* and for \parallel -coords, though it has not been implemented in general. Incidentally, this is a wonderful M.Sc. thesis topic.
6. The real value of visualization, in my opinion, is not seeing “zillions of objects” but rather recognizing *relations* among them. We know that projections lose information that can possibly be recovered with interactivity. Nevertheless, important *clues that can guide the exploration* are lost. So it is preferable to *start* with a display containing all the information, though it may be difficult to uncover. The visual cues available and how they guide the exploration are crucial deciding factors.
7. The value of rigor is self-evident.

These and additional issues that constitute the discovery process are better appreciated via the exploration of real data sets. The basic queries are introduced first with the example of a satellite and then a much larger astronomical data set. Subsequently, they are combined with Boolean operators to form complex queries applied to financial data. An example with several hundred variables is discussed briefly before we move to automatic classification. Visualization and \parallel -coords play key roles in the geometric algorithm’s conception, internal function, and visual presentation of the classification rule. The minimal set of variables needed to state the rule is found and ordered according to their predictive value.

Mathematical background is interspersed to provide a deeper understanding and wiser use of \parallel -coords and its applications. Specifically,

1. learning the *patterns* corresponding to the basic relations and seeking them out for exploratory data analysis,
2. understanding the design and use of *queries*,
3. motivating further sophisticated applications to statistics such as *response surfaces* [67], and
4. applications to *regression*; see the chapter on proximate planes,
5. understanding that the relational information resides in the *crossings*,
6. *concentrating the relational information in the data in clear patterns eliminating the polygonal lines altogether* as with the “proximate planes” is feasible.

Before entering into the nitty gritty, we pose a visualization challenge that we ask the reader to ponder. For a plane

$$\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0, \quad (10.1)$$

allow the coefficients to vary within a small interval. This generates a family of “close” (let us call them *proximate*) planes

$$\Pi = \{\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0, \quad c_i \in [c_i^-, c_i^+], \quad i = 0, 1, 2, 3\}. \quad (10.2)$$

These are the planes generated by small rotations and translations of π with respect to the three coordinate axes. Altogether they form a “twisted slab,” which even in 3-D with *orthogonal axes* is difficult to visualize. Conversely, given many points in 3-D, how can it be discovered, using *any* general visual method you like, that they lie on a twisted slab and how can such a creature be visualized and described precisely, for $N = 3$ and then for *any* N ? In the meantime, you can project, pursue, scale, reduce, scintillate, regress, corollate, or tribulate to your heart’s content, but please do not peek at the answer given at the end.

10.2 Exploratory Data Analysis with ||-coords

10.2.1 Multidimensional Detective

Parallel coordinates transform multivariate relations into 2-D patterns suitable for exploration and analysis. For this reason they are included in many software tools. The queries “parallel coordinates + software” on Google returned about 31,000 “hits” and “scatterplot matrix + software” about 15,000. Irrespective of the apparent 2:1 relative ratio, the comparable numbers for the two astounded me, having heard the appellations “esoteric,” “unnatural,” “difficult,” “squiggly,” and more for ||-coords after their introduction.

The exploration⁵¹ paradigm is that of a *detective*, starting from the data, searching for clues leading to conjectures, testing, backtracking until *voilà*, the “culprit” is discovered! The task is especially intricate when many variables (i.e., dimensions) are involved, calling for the employment of a *multidimensional* detective (MD). As if there were any doubt, our display of choice is \parallel -coords, where the data appears in the by now familiar squiggly blotches and which, by means of *queries*, the MD skilfully dissects to find precious hidden secrets.

During the ensuing interaction, think, dear reader, how similar queries can be stated using other exploration methodologies, including the ubiquitous spreadsheets. More important, what visual clues are available that would *prompt* the use of such queries. A few basics (covered in Chapter 3) are recalled. In \parallel -coords due to the *point* \leftrightarrow *line* and other dualities, some but *not* all actions are best performed in the dual. The queries, which are the “cutting tools,” operate on the display, i.e., *dual*. Their design should exploit the methodology’s strengths and avoid its weaknesses, rather than mimic the action of queries operating on standard “nondual” displays. Like a surgeon’s many specialized cutting tools, one of our early software versions had many specialized queries. Not only was it hard to classify and remember them but they still could not handle all situations encountered. After experimentation, I chose a few (three) intuitive queries called *atomic*, which can be combined via *Boolean* operations to form complex intricate cuts. Even for relatively small data sets, the \parallel -coords display can look uninformative and intimidating. Lack of understanding of the basics of the underlying geometry and poor choice of queries limits the use of \parallel -coords to unrealistically small data sets. Summarizing, the requirements for successful exploratory data analysis are:

- an informative data display,
- good choice of queries, and
- skillful interaction with the display.

10.2.2 An Easy Case Study: Satellite Data

The first admonition is

- *do not let the picture intimidate you,*

as can easily happen by taking an uninformed look at Fig. 10.2 showing the dataset to be explored. It consists of over 9,000 measurements with 9 variables, the first

⁵¹The venerable name “exploratory data analysis” (EDA) is used interchangeably with the currently more fashionable “visual data mining.”

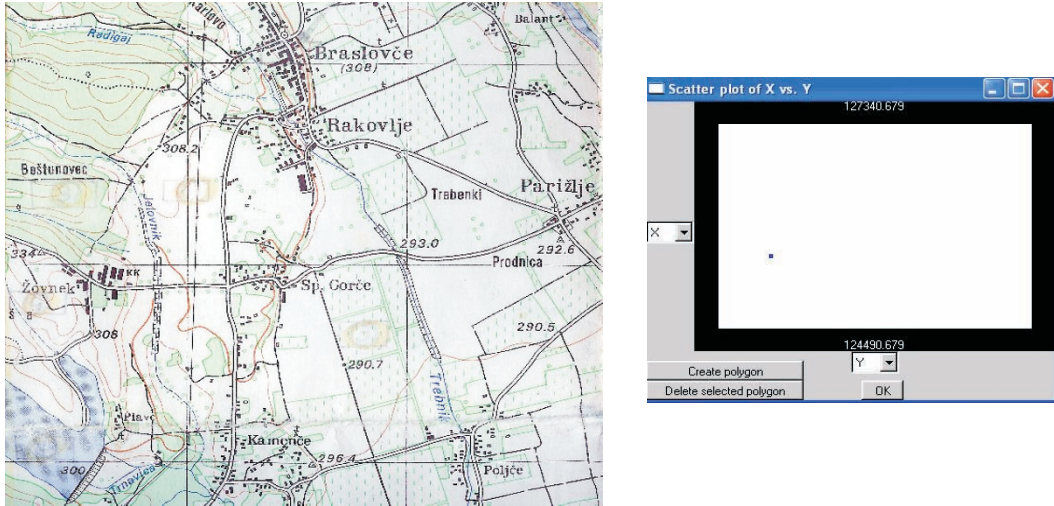


Figure 10.1. Seven types of ground emissions were measured in this region of Slovenia. Measurements recorded by the LandSat Thematic Mapper are shown in subsequent figures. Thanks to Dr. Ana Tretjak and Dr. Niko Schlamberger, Statistics Office of Slovenia, for providing the data. (Right) The display is the map's rectangular region; the dot marks the position where the 7-tuple shown in the next figure was measured.

two (X, Y) specify the location on the map in Fig. 10.1 (left), a portion of Slovenia, where 7 types of ground emissions are measured by satellite. The ground location, (X, Y) , of one data item is shown in Fig. 10.1 (right), which corresponds to the map's region and remains open during the exploration. The query, shown in Fig. 10.2, used to select the data item is called *Pinch*. It is activated by the button **P** on the tool bar. By means of this query, a set of polygonal lines (i.e., data items) can be chosen by being “pinched” *between* the axes. The cursor's movement changes the position of the *selected* arrowhead, which is the larger of the two shown. In due course, various parts of the *GUI* are illustrated (*Parallax*).⁵² Aside from starting the exploration without biases, it is essential

- *to understand the objectives.*

Here the task is the detection and location of various ground features (built-up areas, vegetation, water, etc.) on the map. There is a prominent lake in the lower-left corner with an unusual shape of an upward-pointing “finger.” This brings up the next admonition, that no matter how messy it looks,

⁵²MDG's Ltd proprietary software; all rights reserved, is used by permission.

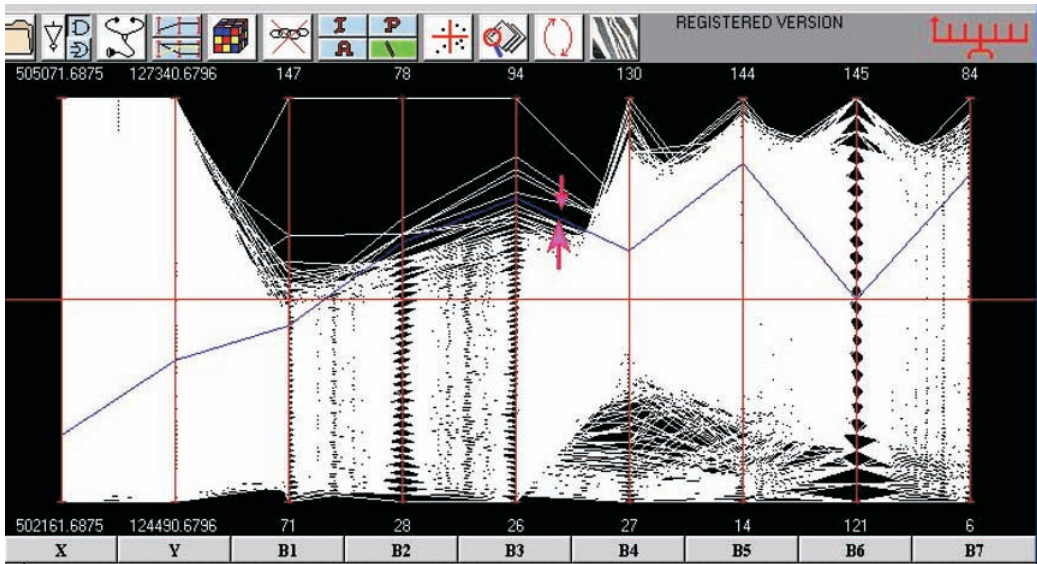


Figure 10.2. Query on parallax showing a single data item.

The X , Y (position, also shown on the right of Fig. 10.1), and values of the 7-tuple ($B1$, $B2$, $B3$, $B4$, $B5$, $B6$, $B7$) at that point.

- *carefully scrutinize the data display for clues and patterns.*

Follow up on anything that catches the eye: gaps, regularities, holes, twists, peaks and valleys, density contrasts like the one at the lower values of $B3$ through $B7$. Using the *Interval* query, activated by the **I** button, starting at the minimum we grab the low range of $B4$ (between the arrowheads) stopping at the dense part as shown in Fig. 10.3. The result, on the left of Fig. 10.4, is amazing. We found the water! The lake is clearly visible together with two other regions, which in the map turn out to be small streams. Our scrutiny having been rewarded, we recall the adage,

- *a good thing may be worth repeating.*

Now examining for density variations *within the selected lower interval of $B4$* , we notice another. The lowest part is much denser. Experimenting a bit, appreciating the importance of interactivity, we select the sparse portion, Fig. 10.5, which defines the water's edge (right) Fig. 10.4 and in fact more. By dropping the lower arrow, we see the lake filling up, starting from the edge, i.e., shallow water first. So the lower values of $B4$ reveal the water, and the lowest "measure," the water's depth; not bad for few minutes of playing around. But all this pertains to a single variable, when we are supposed to be demonstrating *multivariate* exploration. This is a valid point,

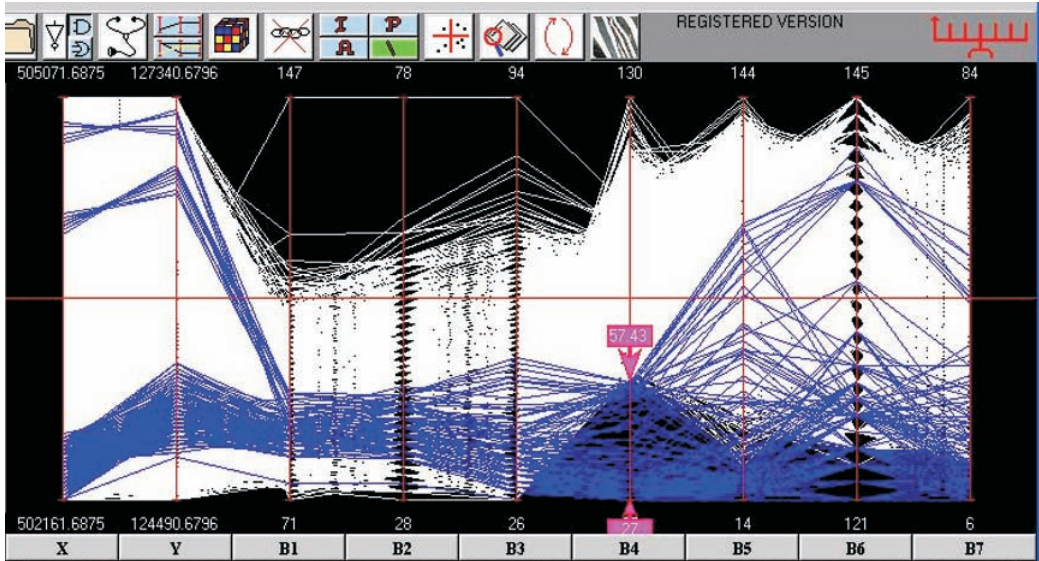


Figure 10.3. Finding water regions.

The contrast due to density differences around the lower values of $B4$ is the *visual cue* that prompts this query.

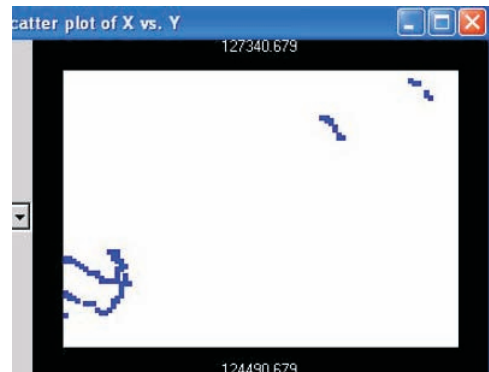
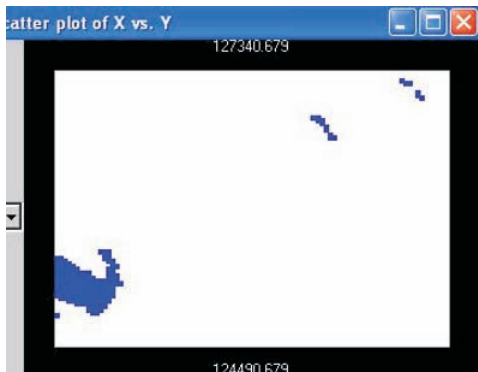


Figure 10.4. (Left) The lake; result of query shown in Fig. 10.3.

On the right is just the lake's edge. It is the result of the query shown in Fig. 10.5.

but we did *pick* $B4$ among several variables. Further, this is a nice “warm-up” for the subsequent more involved examples, enabling us to show two of the queries. The astute observer must have already noticed the regularity, the vertical bands, between the $B1$, $B2$, and $B3$ axes. This is where the *angle* query, activated by the

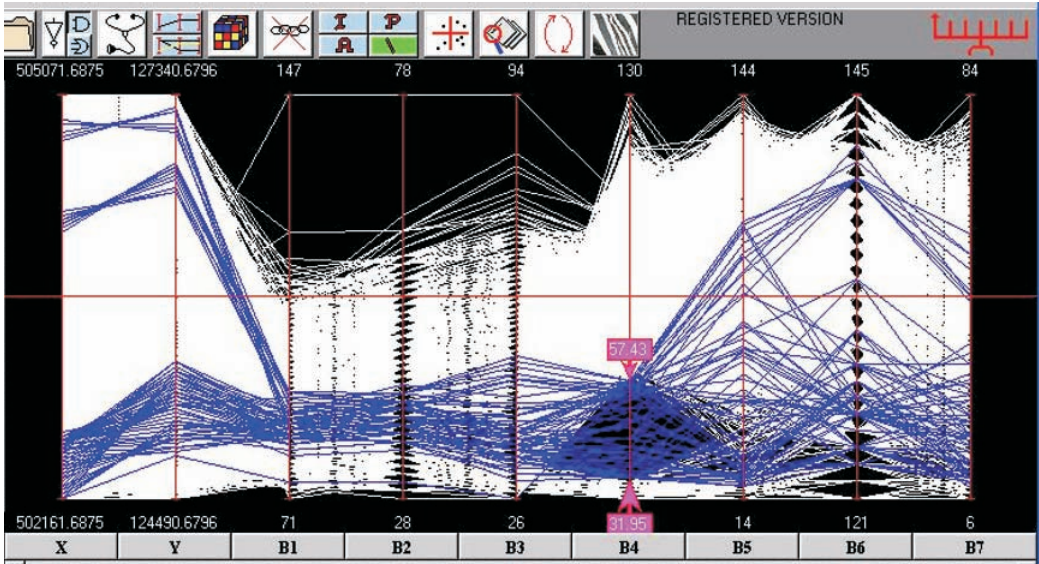


Figure 10.5. Query finding the water's edge.

A button, comes into play. As the name implies, it selects groups of lines within a user-specified angle range. A data subset is selected between the $B2$ and $B3$ axes as shown, with enlarged interaxis distance better showing the vertical bands, in Fig. 10.6 (left) to select a data subset that corresponds on the map to regions with high vegetation. Clicking the **A** button and placing the cursor in the middle of one axis opens an angle, with vertex on the midrange of the previous (left) axis, whose range is controlled by the arrow movements on the right axis. Actually, this “rule” (i.e., relation among some parameters) for finding vegetation can be refined by tweaking a couple of more parameters. This raises the topic of rule finding in general, *classificatio* , which is taken up in Section 10.3.

The *angle* and *pinch* queries are motivated by the ℓ line \rightarrow point $\bar{\ell}$ duality

$$\ell : x_2 = mx_1 + b \leftrightarrow \bar{\ell} = \left(\frac{d}{1-m}, \frac{b}{1-m} \right) \quad (10.3)$$

in ||-coords, illustrated in Fig. 10.7, where the interaxis distance is d . As seen from its x coordinate, the point $\bar{\ell}$ lies between the parallel axes when the line's slope m is negative, to the right of the \bar{X}_2 axis for $0 < m < 1$, and to the left of \bar{X}_1 for $m > 1$. Lines with $m = 1$ are mapped to the *direction* with slope b/d in the xy plane, with d the interaxis distance and b the constant (intercept) in the equation of ℓ . This points out that dualities properly reside in the *projective* plane, the *directions*

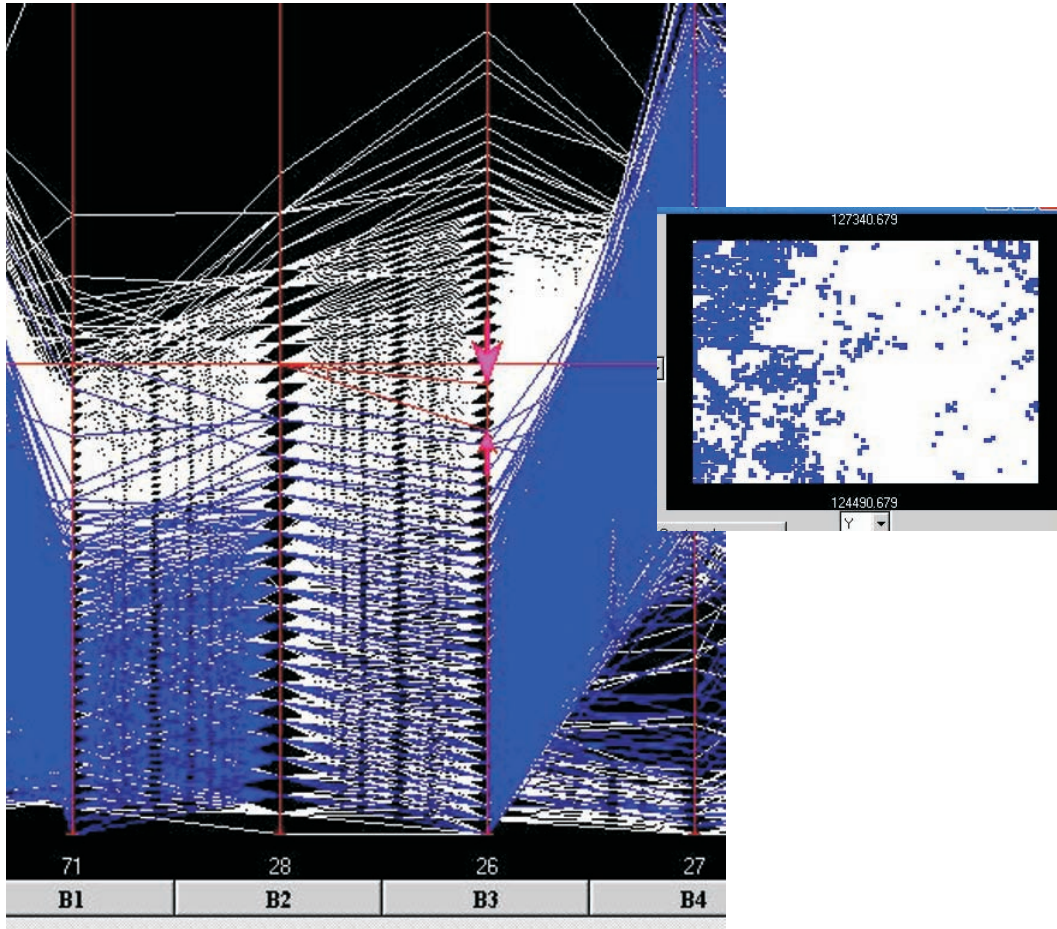


Figure 10.6. Finding regions with vegetation.

Here the *angle query* is used (left) between the *B2* and *B3* axes. Note the arrowheads on the *B3* axis, which specify the angle range for the selected lines.

being the *ideal points*, rather than the Euclidean plane. For sets of points having a “overall” direction with negative slope, i.e., are “negatively correlated,” the lines representing them in \parallel -coords cross each other roughly between the axes, and they can be *selected with the pinch query*. For positively correlated sets of points, their corresponding lines cross outside the axes and can be *selected with the angle query*. All this exemplifies the need to understand some of the basic geometry so as to work effectively with the queries and of course design them properly. The three atomic queries having been introduced, there remains to learn how they can be combined to construct complex queries.

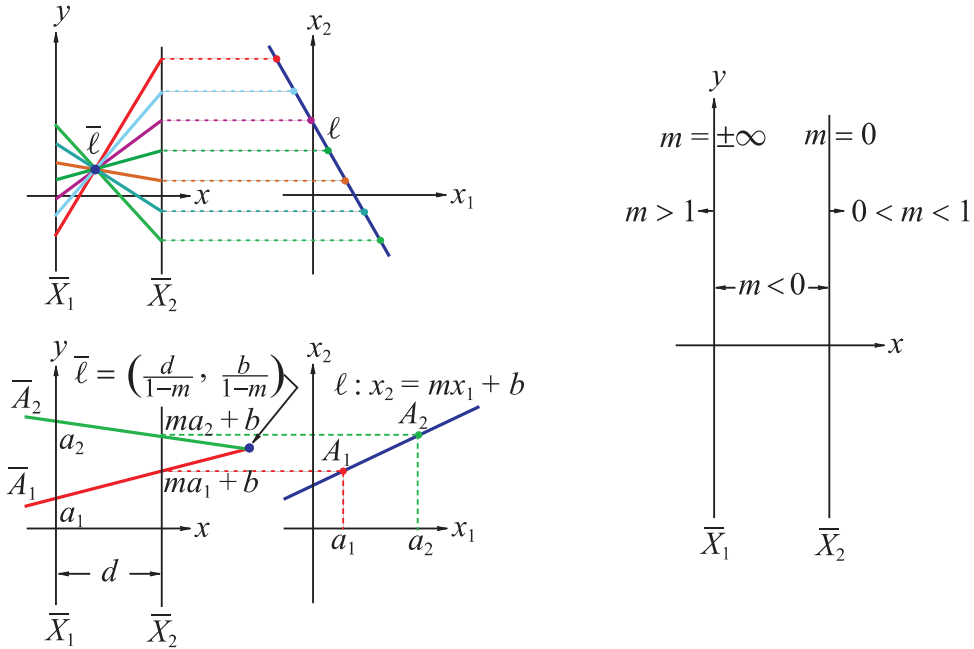


Figure 10.7. Parallel coordinates induce a $point \bar{\ell} \leftrightarrow \ell$ line duality (left).

(Right) The horizontal position of the point $\bar{\ell}$ representing the line ℓ is determined only by the line's slope m . The vertical line $\ell : x_1 = a_1$ is represented by the point $\bar{\ell}$ at the value a_1 on the \bar{X}_1 axis.

Prior to that, Fig. 10.6 (left) raises the question, what if the $B2$ and $B3$ axes were *not* adjacent? Then the pattern and hence their pairwise relation would be missed. Hence the axis permutation used for the exploration is important. In particular, what is the minimum number of permutations among N axes containing the *adjacencies* for all pairs of axes? It turns out [95], [184] that M permutations are needed for even $N = 2M$, and $M + 1$ for odd $N = 2M + 1$. It is fun to see why. Label the N vertices of a graph with the index of the variables $X_i, i = 1, \dots, N$, as shown in Fig. 10.8 for $N = 6$. An edge joining vertex i with j signifies that the axes indexed by i, j are adjacent. The graph on the left is a *Hamiltonian path*, for it contains all the vertices. Such paths have been studied beginning with Euler in the eighteenth century with modern applications to the “traveling salesman” problem and elsewhere [77, pp. 66], [16, pp. 12]. The graph corresponds to the axis index permutation 126354. On the right, the union with the additional two Hamiltonian paths, starting at vertices 2 and 3, forms the complete graph, which contains all possible edges. Hence the three permutations 126354, 231465, 342516 contain all possible adjacent pairs; just try it. The remaining permutations are obtained from the first by successively adding 1 mod 6, and this works for general N .

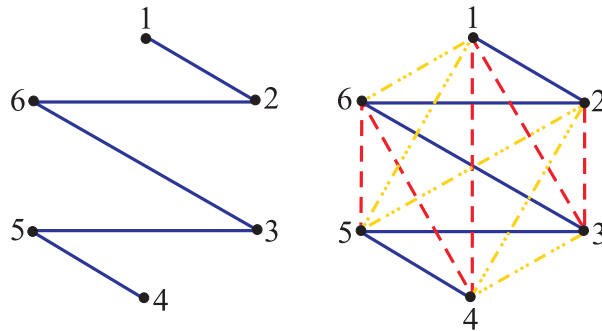


Figure 10.8. (Left) First Hamiltonian path on vertices $1, \dots, 6$.

It corresponds to the (axes) index permutation 126354. (Right) The complete graph as the union of the three distinct Hamiltonian paths starting successively at the vertices 1, 2, 3.

Returning to EDA, the icon with the *Rubik's Cube* on *Parallax's* toolbar activates a *permutation editor*, which automatically generates the Hamiltonian permutations (*HP*). After scrutinizing the data set display, the recommended next step is to run through the $O(N/2)$ *HP*. This is how all nice adjacencies such as the one in Fig. 10.6 are discovered. Then using the editor, patch your own custom-made permutation containing all the parts you like in the *HP*. With this preprocessing cost, referred to earlier in list item 2 of the introduction, the user sets her own best permutation to work with. Of course, there is nothing to prevent one from including axes several times in different positions and experimenting with different permutations in the course of the exploration.

10.2.3 Quasars Dataset

Much larger datasets⁵³ can be explored, and to make this point convincingly, an example from the Quasars catalog having more than 46,000 data items and 21 parameters as displayed in Fig. 10.9 is used.

There are two clusters and a prominent outlier with high value in the second variable. It is evident that several parameters behave similarly and can be omitted, at least initially, uncluttering the display. The reduced dataset is shown on the left of Fig. 10.10, with the two clusters colored differently, enhancing the differences and similarities between them. The *Pinch* query used here is indicated with the opposing arrows; recall that it operates *between the axes*, as contrasted with that in Fig. 10.4, called *Interval*, which operates on a single-variable axis.

⁵³I am grateful to Don Schneider, Jogesh Babu, and Erik Feigelson, who suggested this data set, and Hyunsook Lee, who downloaded the data set from http://astrostatistics.psu.edu/datasets/SDSS_quasar.dat, and assisted with the exploration.

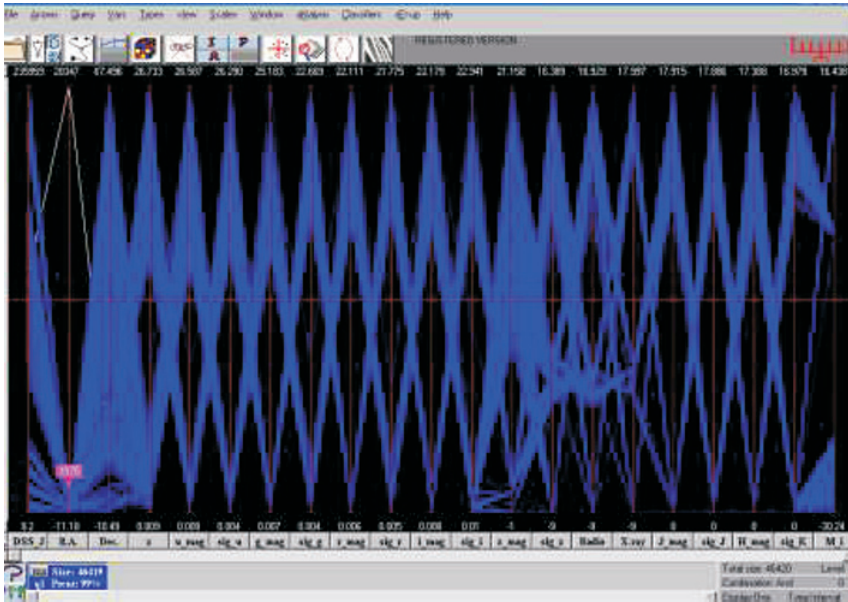


Figure 10.9. The full Quasars dataset with 46, 000⁺ data items and 21 parameters. There are two major clusters and an outlier with high value in the second variable. Several parameters having similar values can be eliminated, uncluttering the display, as shown below in Fig. 10.10.

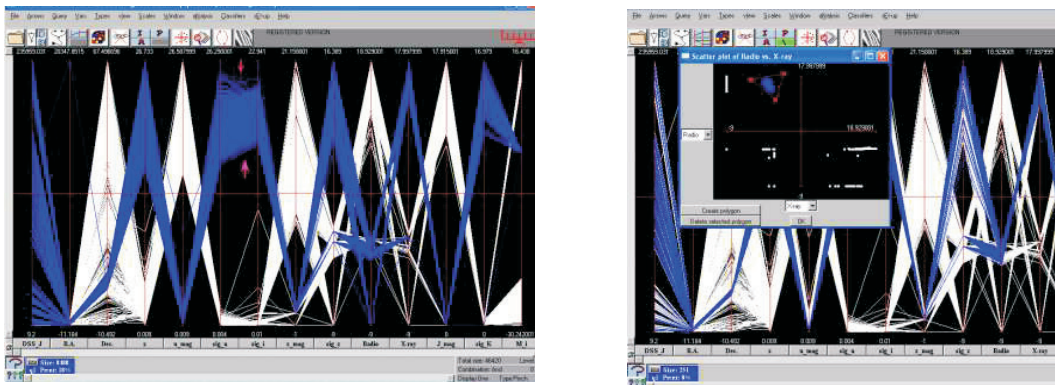


Figure 10.10. The two clusters are colored differently. (Left) The high range between the sixth and seventh variables is selected, revealing patterns involving the intermediate range of the variable labeled *Radio* (appearing as fifth from the right). (Right) On the scatterplot of *X-Ray* vs. *Radio*, the *only* cluster without vertical or horizontal segments is selected.

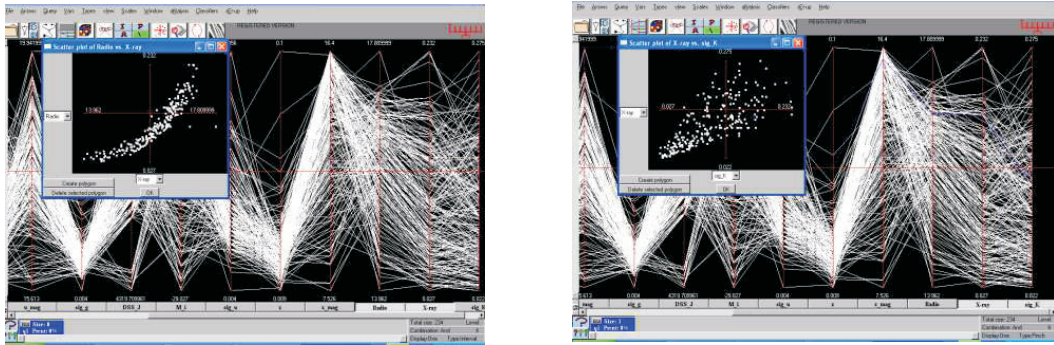


Figure 10.11. The selected cluster consists of 234 data items.
(Left) Relation between *X-Ray* and *Radio*. (Right) Relation between *Radio* and *sig-K*.

Note the patterns involving the intermediate range of values of the parameter labeled *Radio*. On the right within the scatterplot of the variables *Radio* vs. *X-Ray*, the *only* subset without horizontal or vertical segments (for missing values) is selected using a query called *Polygon*. It has 234 data items with interesting relations, as shown in Fig. 10.11, between *Radio* vs. *X-Ray* and *Radio* vs. *sig-K*. The explorations here are relatively straightforward. Much more complex multivariate relations were discovered in complex datasets, and some are shown next. For such a discovery process, more queries are needed whose design requires a deeper understanding of the methodology's foundations. They are summarized next, preparing the ground for some sophisticated applications described later.

10.2.4 Compound Queries: Financial Data

To be explored next is the financial data set shown in Fig. 10.12, the goal being to discover relations useful for investments and trading. The data for the years 1986 (second tick on the third axis) and 1992 are selected and compared. In 1986, the *yen* had the greatest volatility among the three currencies, interests varied in the midrange, gold had a price gap, while *SP500* was uniformly low. By comparison in 1992, the *yen* was stable while the pound sterling was very volatile (possibly due to Soros's speculation); interests and gold prices were low, and the *SP500* was uniformly high. Two interval queries are combined with the *OR* Boolean operator (i.e., union) to obtain this picture. We continue

- “looking for the gold” by checking out patterns that caught our attention.

The data for 1986 is isolated in Fig. 10.13, and the lower range in the gold price gap is selected. Gold prices were low until the second week in August, when they

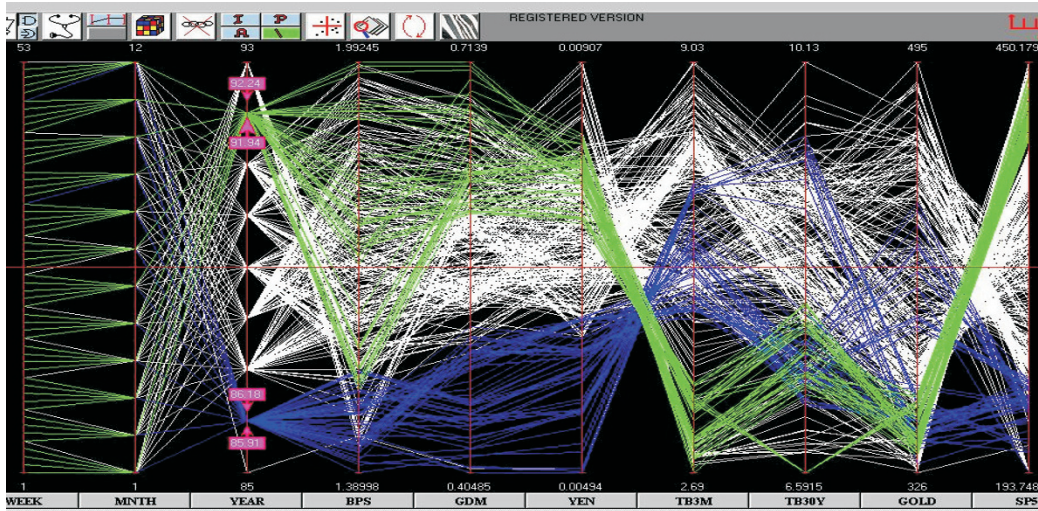


Figure 10.12. Financial data.

Quotes by week on Monday, month, year, first three axes fix the date; Sterling (BPS) D mark, yen rates per dollar; fourth, fifth, sixth axes; 3MTB, 30YTB interest rates in percent for the first three months and 30-year bonds, respectively; seventh, eighth axes; gold in \$/ounce, ninth, SP500 index values on tenth axis.

jumped and stayed higher. The exploration was carried out in the presence of four financial experts, who carefully recorded the relation between low yen, high 3MTB rates, and low gold prices. By the way, the *low yen* rate of exchange means that the yen had a high value relative to the US dollar.

There are two bunches of crossing lines between sixth and seventh axes in Fig. 10.12, which together comprise more than 80% of the data set. This and recalling the previous discussion on the *line* \leftrightarrow *point* mapping in Fig. 10.7 points out the strong negative correlation between yen and 3MTB rates. The smaller cluster in Fig. 10.14 is selected. Moving from the top range of any of the two axes, with the *I* query, and lowering the range causes the other variable's range to rise and is a nice way to show negative correlation interactively. For the contrarians among us, we check also for positive correlation, Fig. 10.15. We find that it exists when gold prices are low to midrange, as happened for a period in the 1990s. This is a free investment tip for bucking the main trend shown in Fig. 10.14. It is also a nice opportunity for showing the *inversion* feature activated by the icon with two cyclical arrows. A variable is selected and the min/max values on that axis are inverted. Diverging lines (as for positive correlation) now intersect Fig. 10.16, making it easier visually to spot the crossing and hence the correlation. Actually,

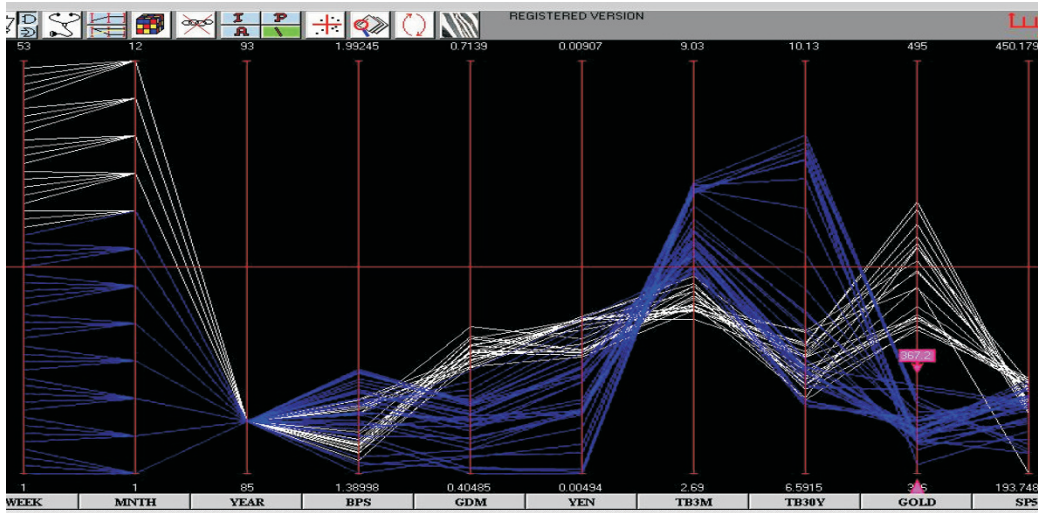


Figure 10.13. Gold prices in 1986.

Gold prices jumped in the second week of August. Note the correlation between the low yen, high 3MTB rates, and low gold price range.

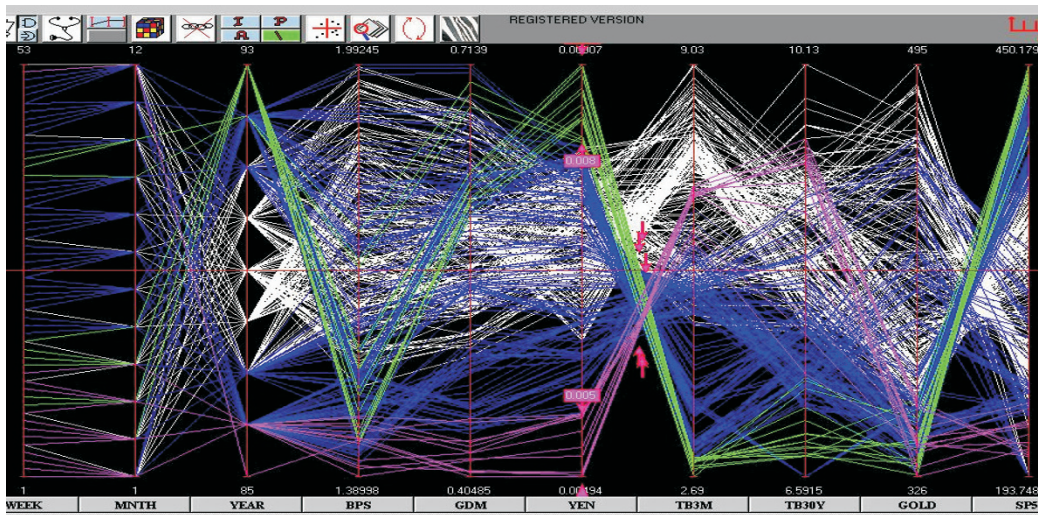


Figure 10.14. Negative correlation.

The crossing lines between the sixth and seventh axes in Fig. 10.12 show strong negative correlation between yen and 3MTB rates. One cluster is selected with the *Pinch* query and combined with the high and low ranges on the yen axis. Data for the years 1986 and 1992 are selected.

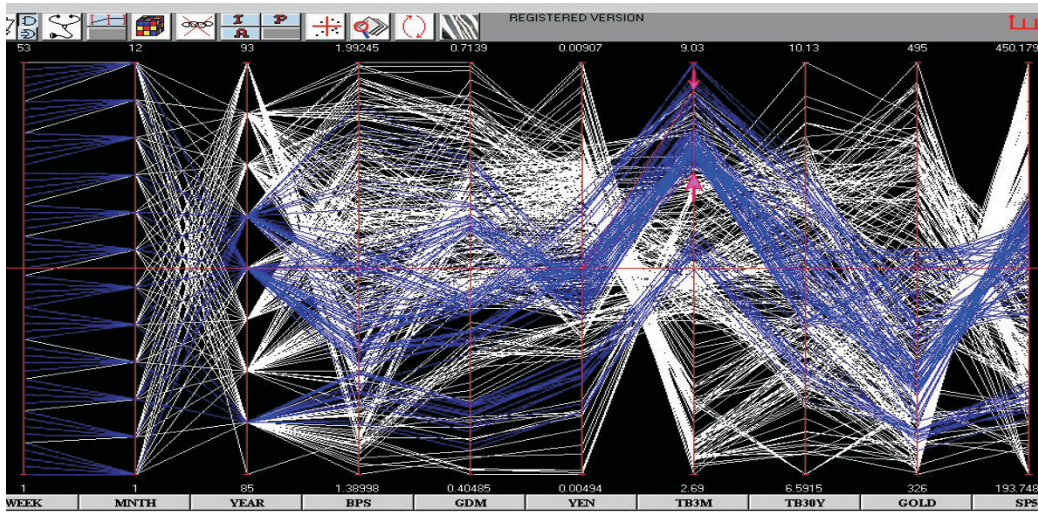


Figure 10.15. Positive correlation.

A positive correlation where the yen and 3MTB rates move together when gold prices are low to midrange.

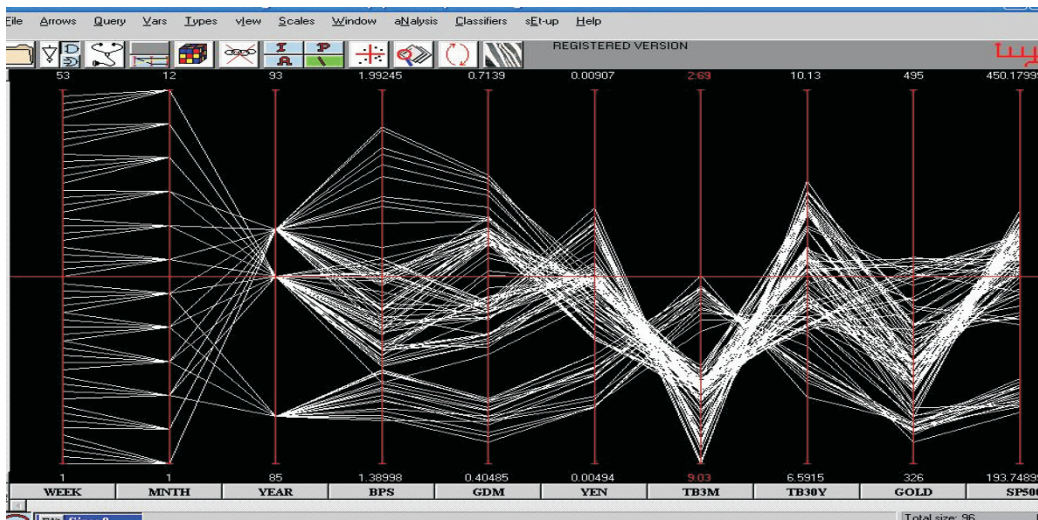


Figure 10.16. Inverting the 3MTB axis.

Now the lines between the yen-3MTB and 3MTB-30MTB axes in Fig. 10.15 cross.

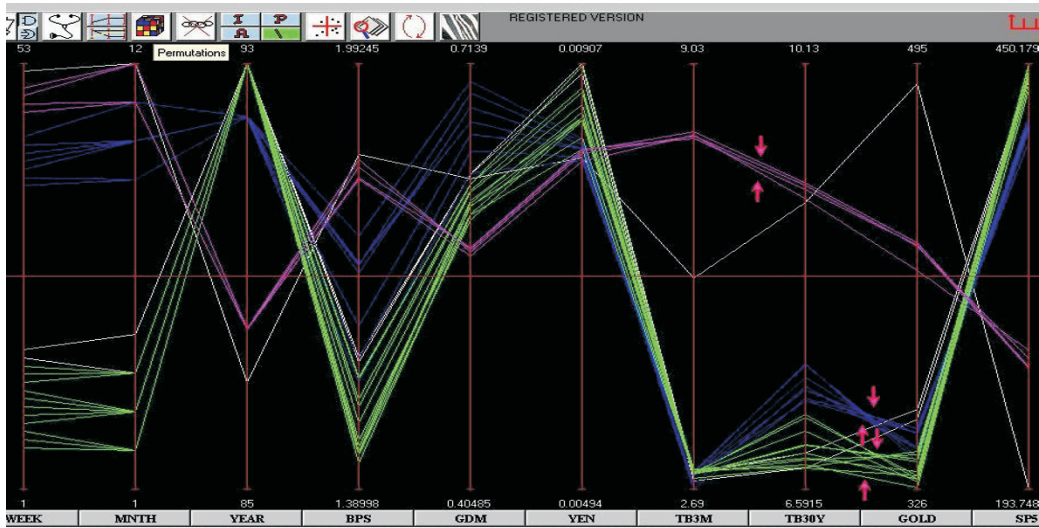


Figure 10.17. Variations in currency exchange rates.

Variations in the rate of exchange of the currencies correlate with movements in the price of gold.

the recommendation is to work with the A query, experimenting with various angle ranges using inversion to check out or confirm special clusters. When stuck, don't just stand there, but

- *vary one of the variables, watching for interesting variations in the other variables.*

Doing this on the *yen* axis, Fig. 10.17, we strike another gold connection. The (rough) intersection of a bunch of lines joining *yen* to the *Dmark* corresponds, by the duality, to their rate of exchange. When the rate of exchange changes, so does the intersection *and the price of gold*! That is, movements in currency exchange rates and the price range of gold go together. Are there any indications that are associated with the high range of gold? The top price range is selected, Fig. 10.18, and prompted by the result of the previous query, we check out the exchange rate between Sterling (BPS) and *Dmark* (or *yen*), and the result is stunning: a perfect straight line. The slope *is* the rate of exchange, which is constant when gold tops out. The relation between Sterling (BPS) and the *Dmark* is checked for different price ranges of gold, Fig. 10.19, and the only regularity found is the one straight line above. Aside from the trading guideline it establishes, it suggests “behind-the-scenes manipulation of the gold market.” We could have said that, but we won't. We perish this thought and proceed with the Boolean complement, Fig. 10.20, of an I (or any other) query. Not finding anything, we select a narrow but dense range on

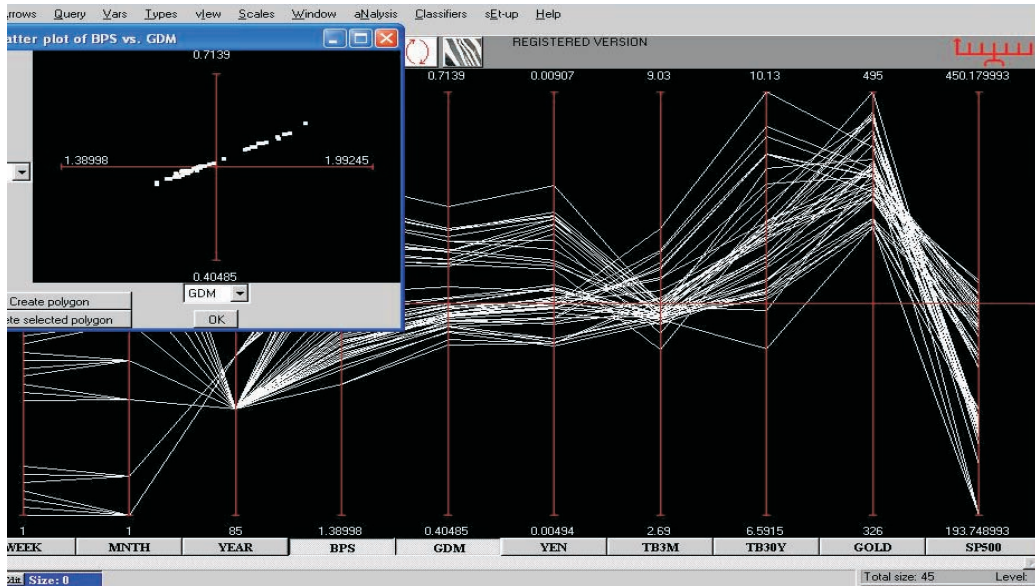


Figure 10.18. High gold.

Note the perfect straight line in the sterling vs. *D*mark plot. The slope is the *rate of exchange* between them, which remains constant when gold prices peak.

the *yen*, Fig. 10.21, and notice an interesting relation between the *D*mark, interest rates, and gold.

There is an exploratory step akin to “multidimensional contouring,” which we fondly call *Zebra*, activated by the last icon button on the right with the appropriate skin color. A variable axis is selected, the *SP500* axis in Fig. 10.22, and divided into a number (user specified) of intervals (here it is 4) and colored differently. This shows the connections (influence) of the intervals with the remaining variables, which here is richly structured especially for the highest range. So what does it take for the *SP500* to rise? This is a good question and helps introduce Parallax’s classifier, described in 10.3. The result, shown in Fig. 10.23, confirms the investment community’s experience that low *3MTB* and gold predict high *SP500*. A comparison with the results obtained on this data set with other visualization tools would be instructive, though unfortunately one is not available. Still let us consider such an analysis done by the scatterplot matrix. There are 10 variables (axes), which requires 45 pairwise scatterplots each, even with a large monitor screen being no larger than about $2.5 \times 2.5 \text{ cm}^2$. Varying one, two, or more variables in tandem and observing the effects *simultaneously* over *all* the variables in the 45 squares may be possible but quite challenging. By contrast, the effects of varying the *D*mark, *conditionally* for

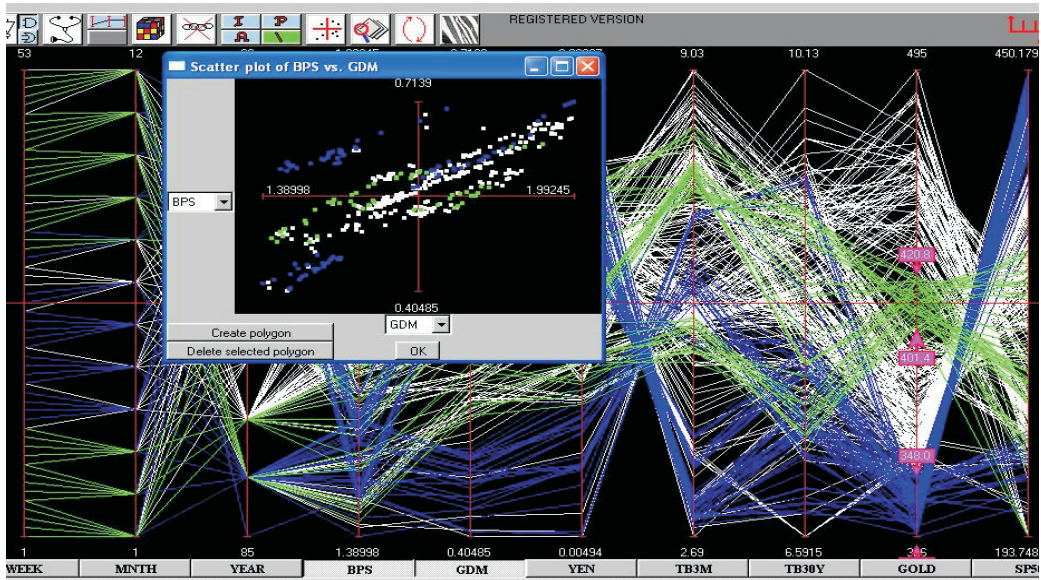


Figure 10.19. Two price ranges of gold.
The associated sterling vs. *D*mark plots show no regularity.

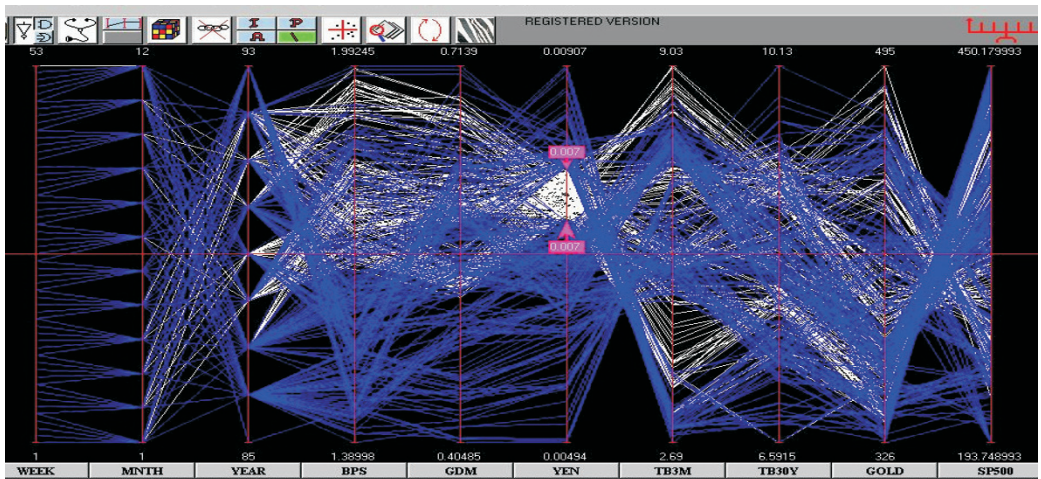


Figure 10.20. The complement of an *I* query.

stable yen, are easily seen on the two interest rates, gold as well as the remaining variables in *one*, Fig. 10.21. This example illustrates the difficulties due to high

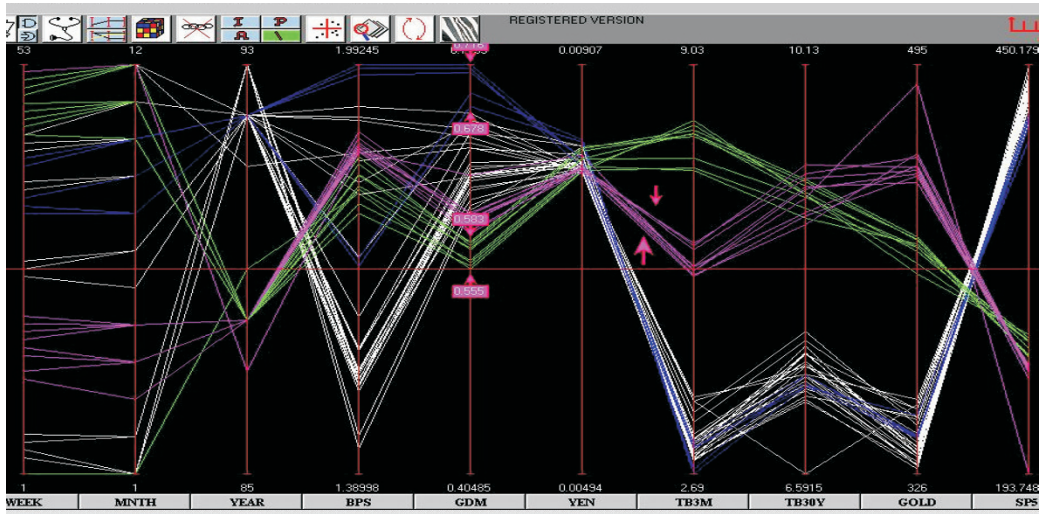


Figure 10.21. *Yen stable.*

For the *yen* trading in a narrow range, high *Dmark* goes with low *3MTB* rates, low *Dmark* goes with high *3MTB* rates, while mid *3MTB* rates go with high gold.

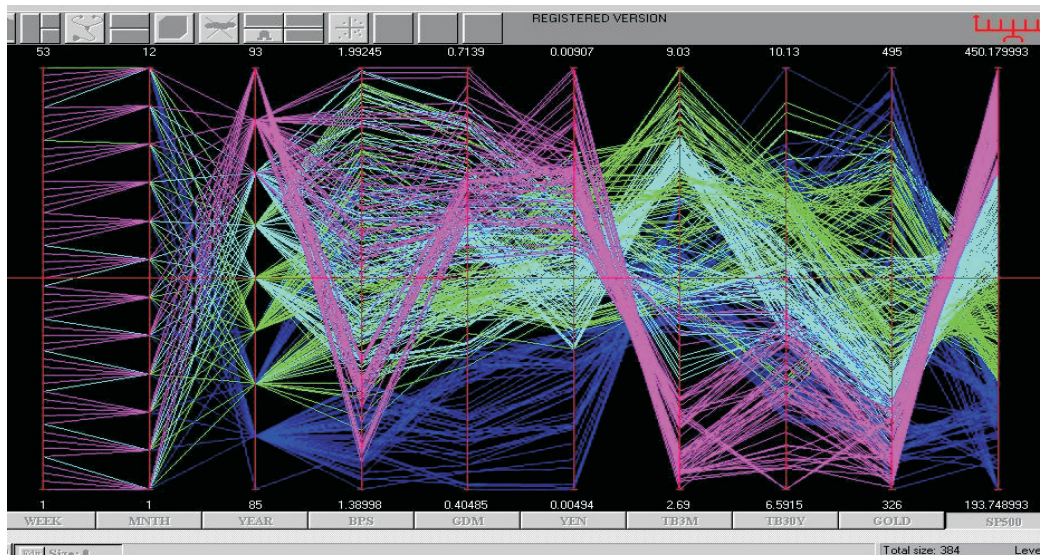


Figure 10.22. *The zebra query.*

It partitions and colors the segments of values differently. A variable, here the *SP500* axis, is divided into equal (here 4) intervals. This quickly reveals interrelationships. Note especially those for the highest *SP500* range and see the next figure.

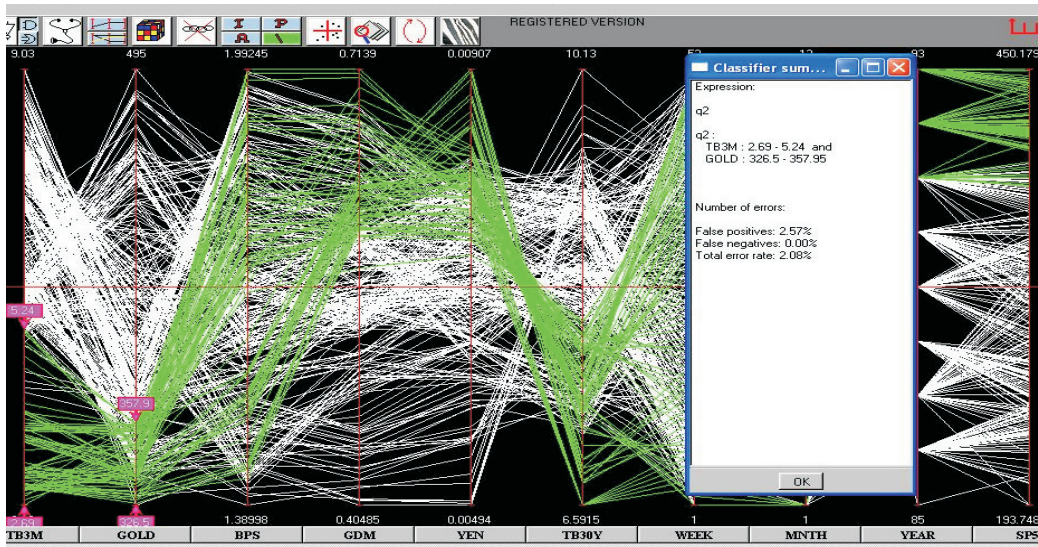


Figure 10.23. The rule for high *SP500*.

Both *3MTB* (the “short-bond” as it is called) and gold are low and in this order of importance.

representational complexity (see Section 10.1.2 item #2), which is $O(N^2)$ for the scatterplot matrix but $O(N)$ for $\|$ -coords and made even clearer with the next data set.

10.2.5 Hundreds of Variables

An important question frequently asked is, how many variables can be handled with $\|$ -coords? The largest data set that I have effectively worked with had about 800 variables and 10,000 data entries. With various techniques developed over the years and the automatic classifier discussed in the next section, much larger data sets can be handled. Still the relevant admonition is

- *be skeptical about the quality of data sets with a large number of variables.*

When hundreds or more variables are involved, it is unlikely that there are many people around who have a good feel for what is happening, as confirmed by my experience. A case in point is the data set shown in Fig. 10.24, consisting of instrumentation measurements for a complex process. The immediate observation is that many instruments recorded 0 for the duration, something which was unnoticed by the process monitors. Another curiosity was the repetitive patterns on the right. It turns that several variables were measured in more than one location using

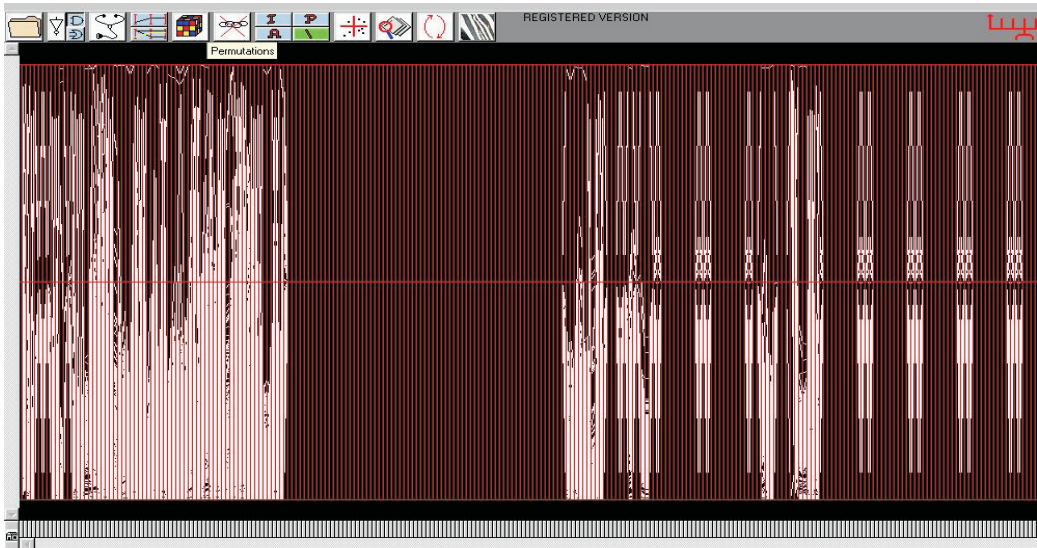


Figure 10.24. Manufacturing process measurements; 400 variables.

different names. When the data set was cleaned up of superfluous information, it was reduced to about 90 variables, as shown in Fig. 10.25 and eventually to about 30, which contained the information of real interest. By my tracking, the phenomenon of repetitive measurements is widespread, with at least 10% of the variables occurring in large data sets being duplicates or near duplicates, possibly due to instrument nonuniformities, as suggested in the 2-variable scatterplot2 in Fig. 10.25. Here the repetitive observations were easily detected due to the fortuitous variable permutation in the display. Since repetitive measurements occur frequently, it may be worth adding to the software an automated feature to detect and exhibit the suspect variables. This brief exposure is just an indication that large (in dimension, i.e., in number of variables) data sets can still be gainfully explored in $\|$ -coords.

There follows a different example of EDA on a process-control data set [102], where compound queries turned out to be very useful and where we learn to add to the list of exploration guidelines arguably the most important one:

- *test the assumptions and especially the “I am really sure of”s.*

10.2.6 Production of VLSI (Chips)

The data set displayed in Fig. 10.26 consists of production data of several batches of a specific VLSI (computer chip) with measurements of 16 parameters involved

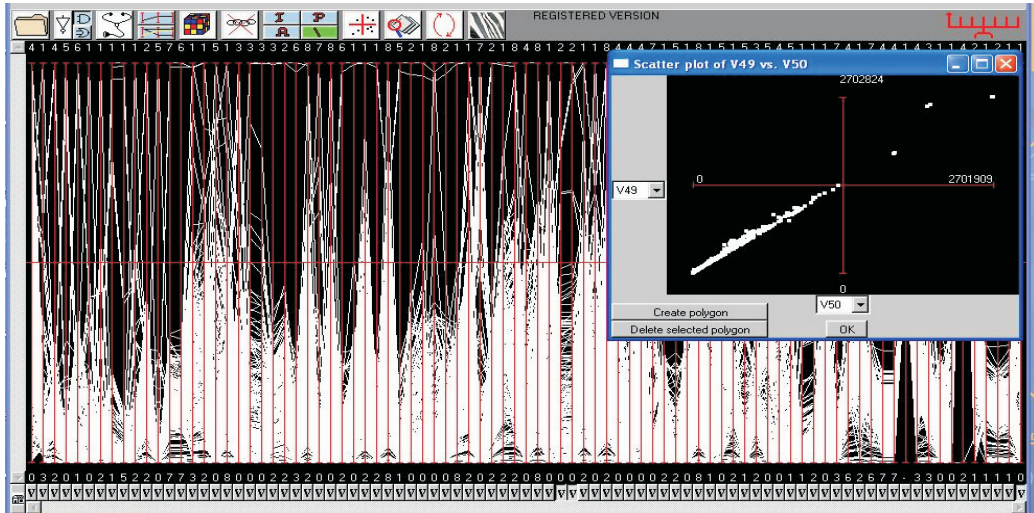


Figure 10.25. The above data set after “cleanup” with about 90 variables left.

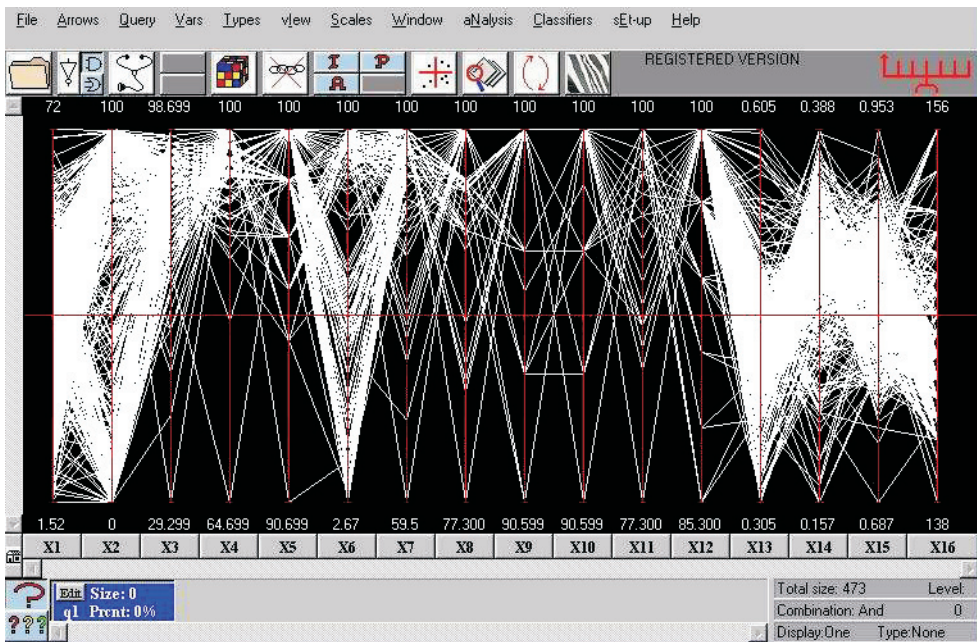


Figure 10.26. Dataset; VLSI production with 16 parameters.

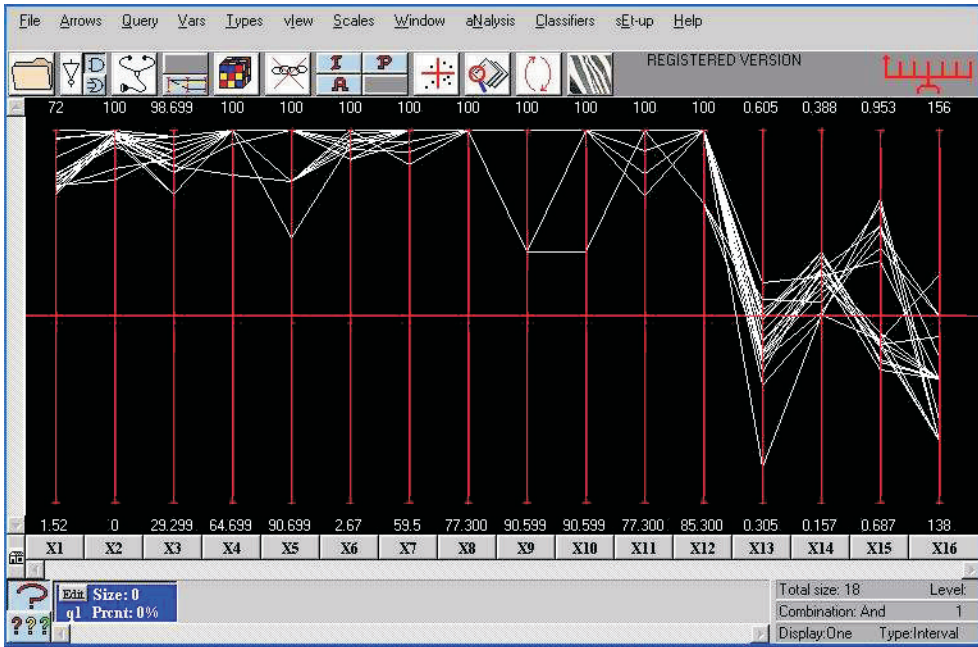


Figure 10.27. The batches high in Yield, X_1 , and Quality, X_2 .

in the process. The parameters are denoted by X_1, X_2, \dots, X_{16} . The *yield*, as the percentage of useful chips produced in the batch, is denoted by X_1 , and X_2 is a measure of the *quality* (given in terms of speed performance) of the batch. Ten different categories of *defects* are monitored, and the variables' scales of X_3 through X_{12} are 100% pure (zero defect) so that the amounts of defect appear at the top and increasing amounts appear proportionately lower. The remaining X_{13} through X_{16} denote some physical parameters.

Since the goal here is to raise the yield, X_1 , while maintaining high quality, X_2 , we have the problem of multiobjective optimization due to the presence of more than one objective. The production specialists believed that it was the presence of defects that prevented high yields and qualities. So their purpose in life was to keep on pushing — at considerable cost and effort — for zero defects.

With this in mind, the result of our first query is shown in Fig. 10.27, where the batches having the highest X_1 and X_2 have been isolated. This is an attempt to obtain clues; and two real good ones came forth. Notice the resulting range of X_{15} , where there is a significant separation into two clusters. As it turns out, this gap yielded important insight into the physics of the problem. The other clue is almost hidden. A careful comparison — and here interactivity of the software is

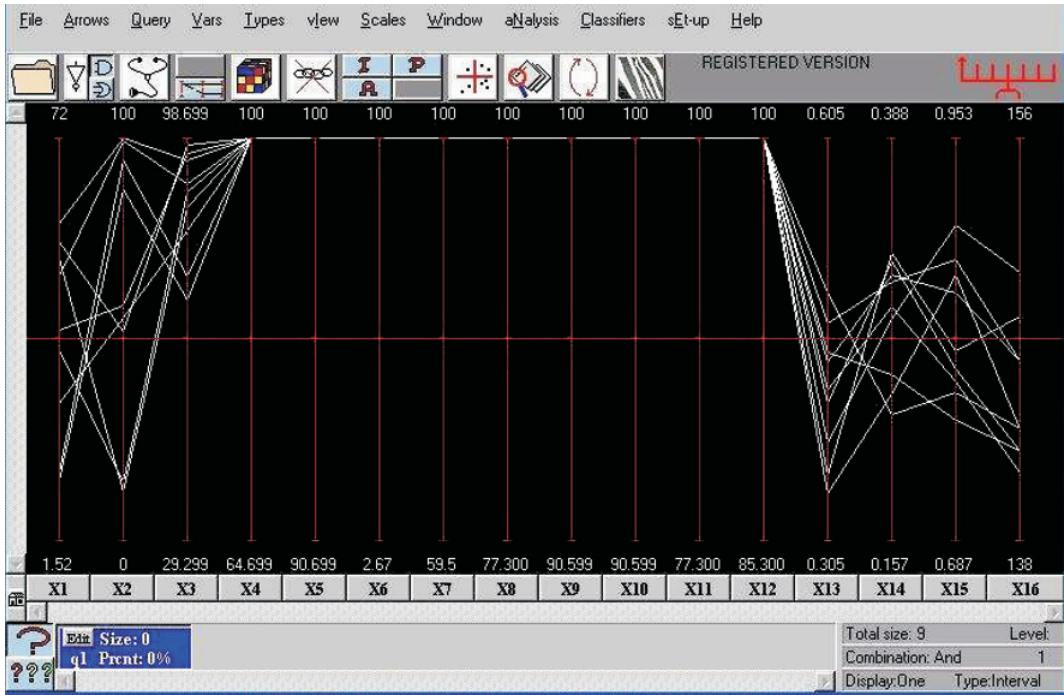


Figure 10.28. The batches with zero in nine out of ten defect types.

essential — between Fig. 10.26 and Fig. 10.27 shows that some batches that were high in X3 (i.e., low in that defect) were not included in the selected subset. That casts some doubt into the belief that zero defects are the panacea and motivates the next query, where we search for batches having zero defects in at least nine (excluding X3, where we saw that there are problems) out of the ten categories. The result is shown in Fig. 10.28 and is a shocker. There are nine such batches, and all of them have poor yields and for the most part also low quality! That this was not questioned and discovered earlier is surprising. We scrutinize the original picture Fig. 10.26 for visual cues relevant to our objectives and our findings so far. And there is one staring us in the face! Among the ten defects X3 through X12, whatever X6 is, its graph is very different from those of the others. It shows that the process is much more sensitive to variations in X6 than the others. For this reason, we chose to treat X6 differently and remove its zero-defect constraint. This query (not shown) showed that the very best batch (i.e., highest yield with very high quality) does not have zeros (or the lowest values) for X3 and X6; a most heretical finding. It was confirmed by the next query, which isolated the cluster of batches with the top yields (note the gap in X1 between them and the remaining batches).



Figure 10.29. The batches with the highest yields.
They do not have the lowest defects of type X3 and X6.

These are shown in Fig. 10.29, and they confirm that small amounts (the ranges can be clearly delimited) of X3- and X6-type defects are essential for high yields and quality.

Returning to the subset of data that best satisfied the objectives, Fig. 10.27, in order to explore the gap in the range of X15, we found that the cluster with the high range of X15 gives the lowest (of the high) yields X1, and worse, it does not give consistently high quality X2, whereas the cluster corresponding to the lower range has the higher qualities and the full range of the high yield. It is evident that the small ranges of X3, X6 close to (but not equal to) zero, together with the short (lower) range of X15, provide necessary conditions for obtaining high yields and quality. This is also indicated in Fig. 10.29. By a stroke of good luck, these three can also be checked early in the process, avoiding the need of “throwing good money after bad” (i.e., by continuing the production of a batch whose values of X3, X6, and X15 are not in the small “good” ranges we have found).

These findings were significant and differed from those found with other methods for statistical process control [9]. This approach has been successfully used

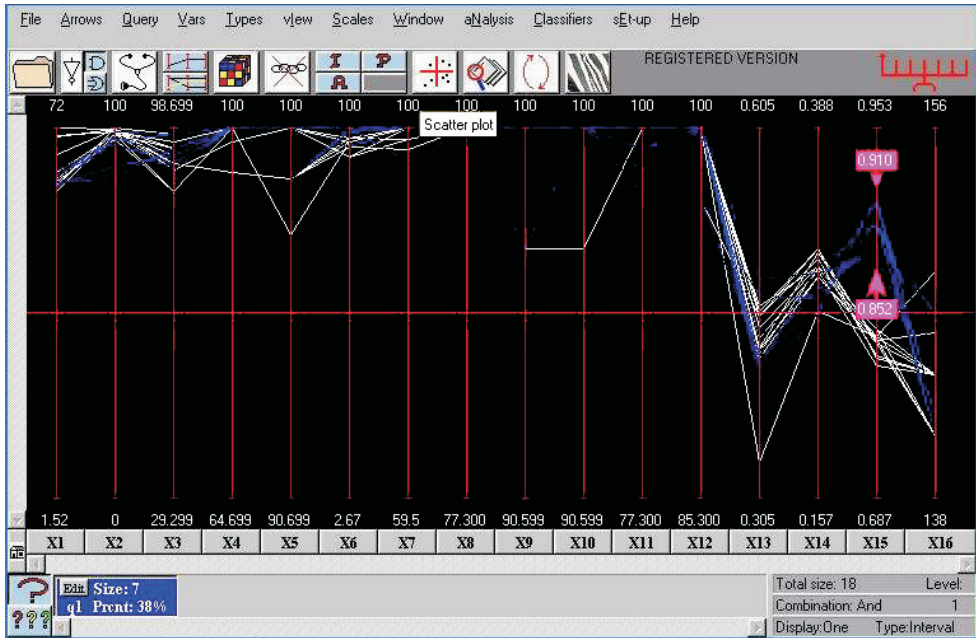


Figure 10.30. Only the lower range of X15 is associated with the highest yields and quality.

in a wide variety of applications, including the manufacture of printed circuit boards, PVC and manganese production, financial data, and determining skill profiles (i.e., as in drivers, pilots).

10.3 Classification

Though it is fun to undertake this kind of exploration, the level of skill and patience required tends to discourage some users. It is not surprising, then, that the most persistent requests and admonitions have been for tools that, at least partially, automate the knowledge discovery process [103]. Classification is a basic task in data analysis and pattern recognition, and an algorithm accomplishing it is called a *classifier* [149], [57], [136]. The input is a dataset P and a designated subset S . The output is a characterization, a set of conditions or rules, to distinguish elements of S from all other members of P , the “global” dataset. The output may also be that there is insufficient information to provide the desired distinction.

With parallel coordinates, a dataset P with N variables is transformed into a set of points in N -dimensional space. In this setting, the designated subset S can be described by means of a hypersurface that encloses just the points of S .

In practical situations, the strict enclosure requirement is dropped, and some points of S may be omitted (“false negatives”), and some points of $P - S$ are allowed (“false positives”) in the hypersurface. The description of such a hypersurface is equivalent to the rule for identifying, within some acceptable error, the elements of S . Casting the problem in a geometric setting leads us to *visualize* how such a rule may work. This entails:

1. Use of an efficient “wrapping” (a convex-hull approximation) algorithm to enclose the points of S in a hypersurface S_1 containing S and in general also some points of $P - S$; so $S \subset S_1$.
 2. The points in $(P - S) \cap S_1$ are isolated and the wrapping algorithm is applied to enclose them, and usually also some points of S_1 , producing a new hypersurface S_2 with $S \supset (S_1 - S_2)$.
 3. The points in S not included in $S_1 - S_2$ are next marked for input to the wrapping algorithm, a new hypersurface S_3 is produced containing these points as well as some other points in $P - (S_1 - S_2)$, resulting in $S \subset (S_1 - S_2) \cup S_3$.
 4. The process is repeated, alternatively producing upper and lower containment bounds for S ; termination occurs when an error criterion (which can be user-specified) is satisfied or when convergence is not achieved. After termination is obtained, two error measures are available to estimate the rule’s precision:
- *Train & Test.* A portion of the dataset (usually 2/3) selected at random is used to derive the classification rule, which is then tested on the remaining 1/3 of the data.
 - *Cross-Correlation.*

It can and does happen that the process does not converge when P does not contain sufficient information to characterize S . It may also happen that S is so “porous” (i.e., spongelike) that an inordinate number of iterations are required. On convergence, say at step $2n$, the description of S is provided as

$$S \approx (S_1 - S_2) \cup (S_3 - S_4) \cup \cdots \cup (S_{2n-1} - S_{2n}), \quad (10.4)$$

this being the terminating expression resulting from the algorithm that we call *nested cavities (NC)*.

The user can select a subset of the available variables and restrict the rule generation to these variables. In certain applications, as in process control, not all variables can be controlled, and hence it is useful to have a rule involving only the accessible (i.e., controllable) variables. An important additional benefit is that the minimal set of variables needed to state the rule is found and ordered according

to their predictive value. These variables may be considered as the best *features* to identify the selected subset. The algorithm is display-independent and there is no inherent limitation as to the size and number of variables in the dataset. Summarizing for *NC*,

- an approximate convex-hull boundary for each cavity is obtained,
- utilizing properties of the representation of multidimensional objects in $\|$ -coords, a very low polynomial worst-case complexity of $O(N^2|P|^2)$ in the number of variables N and dataset size $|P|$ is obtained; it is worth contrasting this with the often unknown, or unstated, or very high (even exponential) complexity of other classifiers,
- an intriguing prospect, due to the low complexity, is that the rule can be derived in near real time, making the classifier adaptive to changing conditions,
- the minimal subset of variables needed for classification is found,
- the rule is given explicitly in terms of conditions on these variables, i.e., included and excluded intervals, and provides a “picture” showing the complex distributions with regions where there are data and “holes” with no data providing insights to the domain experts.

During 1990–1993, Michie, Spiegelhalter, and Taylor [134], on behalf of the ESPRIT program of the European Union, made extensive studies of several classifiers applied to diverse data sets. About 23 different classifiers were applied to about 20 data sets for comparative trials in the *StatLog* project. This was designed to test classification procedures on large-scale commercially important problems in order to determine suitability of the various techniques to industry. The results obtained by *NC* are compared with those obtained by other well-known classifiers used in *Statlog* on two benchmark datasets and shown in the accompanying tables.

1. *Satellite image dataset* has over 6,000 items and *NC*’s classification error was 9%, k-NN was next with 9.4%, the remainder had errors as high as 30%, and one was unable to be classified at all, as shown in Table 10.1.
2. *Vowel recognition data* with about 1,000 data items. The data collection process involves digital sampling of speech with acoustic signal processing, followed by recognition of the phonemes, groups of phonemes, and words. The goal here is a speaker-independent rule based on 10 variables of 11 vowels occurring in various words spoken (recorded and processed) by 15 British male and female speakers. Deterding [40] collected this dataset of vowels, which can be found in the CMU benchmark repository on the Internet. There are 528 entries for training and 462 for testing. Three other types of classifiers were also applied

Table 10.1. Summary of the *Statlog* results and comparison with the *nested cavities* (*NC*) classifier for the satellite image data.

Rank	Classifie	Error rate %	
		Train	Test
1	NC	4.3	9.0
2	k-NN	8.9	9.4
3	LVQ	4.8	10.5
4	DIPOL92	5.1	11.1
5	RBF	11.1	12.1
6	ALLOC80	3.6	13.2
7	IndCART	2.3	13.8
8	CART	7.9	13.8
9	Backprop	11.2	13.9
10	Baytree	2.0	14.7
11	CN2	1.0	15.0
12	C4.5	4.0	15.0
13	NewID	6.7	15.0
14	Cal5	12.5	15.1
15	Quadisc	10.6	15.5
16	AC ²	11.3	15.7
17	SMART	12.3	15.9
18	Cascade	11.2	16.3
19	Logdisc	11.9	16.3
20	Discrim	14.9	17.1
21	Kohonen	10.1	17.9
22	CASTLE	18.6	19.4
23	NaiveBay	30.8	28.7
24	ITrule	Failed	Failed

to this dataset: neural networks and k-NN by Robinson and Fallside [56], and decision trees by Shang and Breiman [158]. Again *NC* was top with 7.9%; next was CART-DB at 10%; the rest were down to 66%, with many unable to provide a classification rule. The results are shown in Table 10.2.

Table 10.2. Summary of classification results for the vowel dataset.

Rank	Classifie	Testing Mode	Test Error Rate %
1	nested cavities (NC)	cross validation	7.9
2	CART-DB	cross validation	10.0
3	nested cavities (NC)	train & test	10.5
4	CART	cross validation	21.8
5	k-NN	train & test	44.0
6	RBF	train & test	46.5
7	multilayer perceptron	train & test	49.4
8	single-layer perceptron	train & test	66.7

3. *A neural-pulse dataset* has interesting and unusual features. There are two classes of neurons whose outputs to stimuli are to be distinguished. They consist of 32 different pulses measured in a monkey's brain (poor thing!). There are 600 samples with 32 variables (the pulses).⁵⁴ Various classification methods were unable to obtain a rule. With *NC*, convergence is obtained requiring only 9 of the 32 parameters for the classification rule for category 1. The resulting ordering shows a striking separation. In Fig. 10.31, the first pair of variables x_1, x_2 in the original order is plotted on the left. On the right, the best pair x_{11}, x_{14} , as chosen by the classifier's ordering speaks for itself. By the way, the discovery of this manually would require constructing a scatterplot matrix with 496 pairs, then carefully inspecting and comparing the individual plots. The implementation provides all the next-best sections, some of which are shown in Fig. 10.31, to aid the visualization of the rule. The data set consists of two "pretzel-like" clusters wrapping closely in 8-D, one enclosing the other, showing that the classifier can actually "carve" highly complex regions with the cavity shown. One can understand why separation of clusters by hyperplanes or nearest-neighbor techniques can fail badly for such data sets. The rule has 4% error, some of which are shown in Fig. 10.32.

The rules are explicit, "visualizable," optimally ordering the minimal set of variables needed to state the rule without loss of information. There are variations that apply in some situations where the *NC* classifier fails, such as the presence of

⁵⁴I am grateful to Prof. R. Coiffman and his group at the Departments of Mathematics and Computer Science at Yale University for giving me this data set.

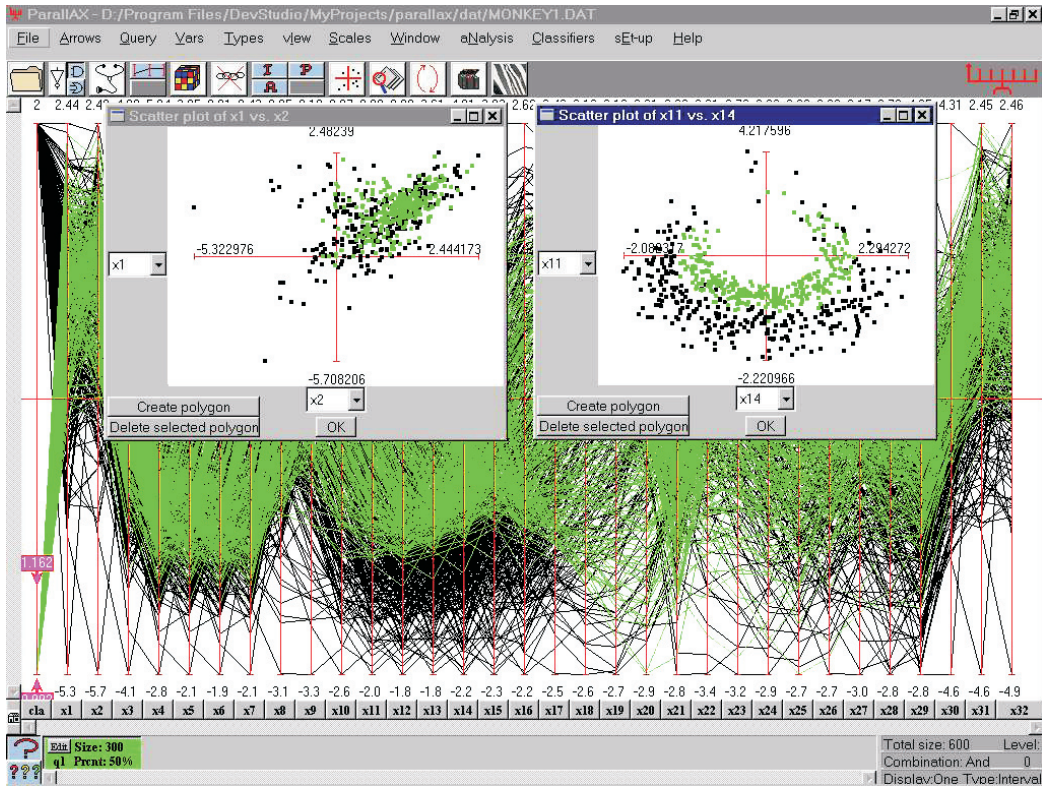


Figure 10.31. The neural-pulses dataset with 32 parameters and two categories. The data set is shown in the background. On the left plot are the first two parameters in the original order. The classifier found the nine parameters needed to state the rule with 4% error and ordered them according to their predictive value. The best two parameters are plotted on the right, showing the separation achieved.

several large “holes” (see [103]). Further, keeping in mind that the classification rule is the result of several iterations suggests heuristics for dealing with the pesky problem of *overfitting*. The iterations can be stopped just where the corrections in (10.4) become very small, i.e., the S_i consist of a small number of points. The number of iterations is user defined, and the resulting rule yields an error in the test stage more stable under variations in the number of points of the test set. In addition, the user can exclude variables from being used in the description of the rule; those ordered last are the ones providing the smaller corrections and hence more liable to overcorrect.

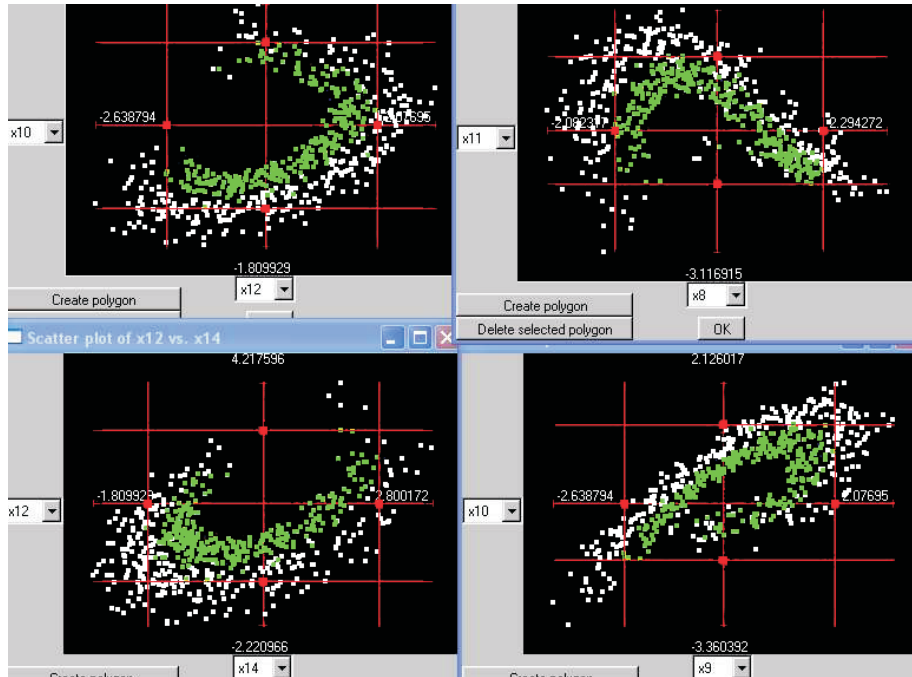


Figure 10.32. Neural dataset classification.

Further cross-sections of the hypersurface corresponding to the classification rule.

10.3.1 Case Study: Noise Signature Recognition

It is worthwhile going through another classification example in greater detail. About 1000 vehicles (trucks) were run on a road test tract embedded with sound sensors at various locations and distances. The resulting noise was measured separately within seven frequency ranges. For each range, the measurements were subsequently transformed (via wavelets) into a single number. The resulting seven numbers v_1, v_2, \dots, v_7 provide the *noise signature* for each specific vehicle. The truck data is divided into 11 classes (by type, manufacturer). Each data item (for a truck) consists of nine parameters: the truck ID, class #, and noise signature. The data set is shown in Fig. 10.33. There are trucks manufactured in one country A and belonging to the army of one country I ; the remainder are made in country R and belong to the army of a hostile country S .

The army of country I would like to detect the enemy's vehicles from a distance by their noise signature (i.e., the way they sound), something analogous to "land sonar." The vehicles of country S are the classes #7 – 11, which are selected and shown in Fig. 10.34. The values of v_4 show that there are two vehicle clusters.

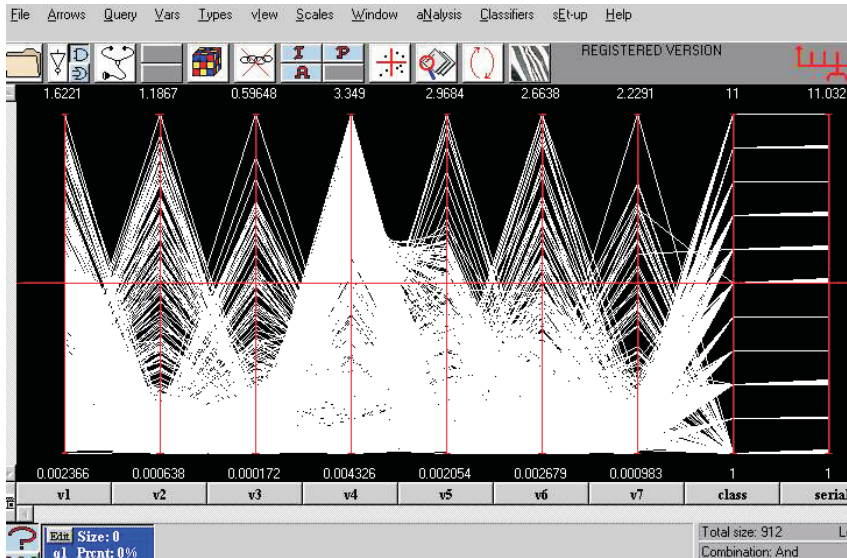


Figure 10.33. Eleven categories of vehicles (trucks).
The *noise signature* defined by the values of v_1, v_2, \dots, v_7 is given for each vehicle.

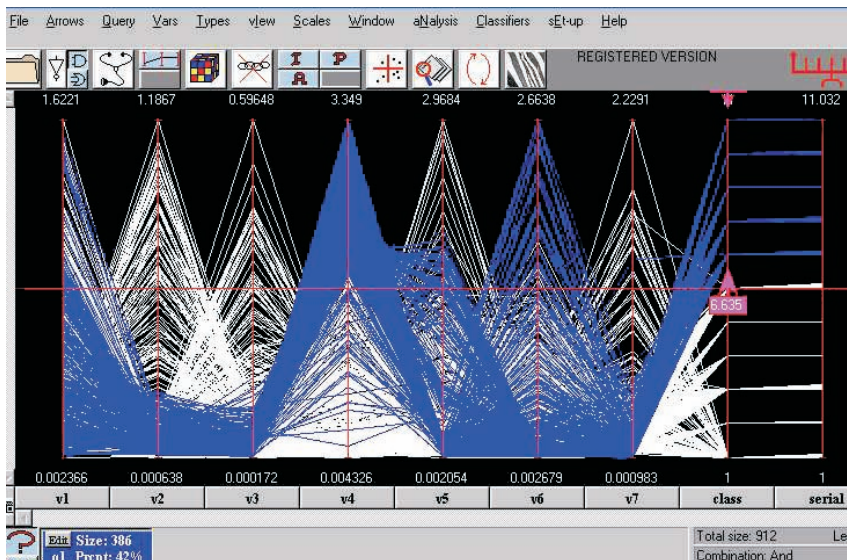


Figure 10.34. Top five classes consist of vehicles manufactured in country *R*.
The values of v_4 indicate two clusters. The values in the top dense range turn out to be for trucks with automatic transmission, while the remaining lower values correspond to trucks with manual transmission.

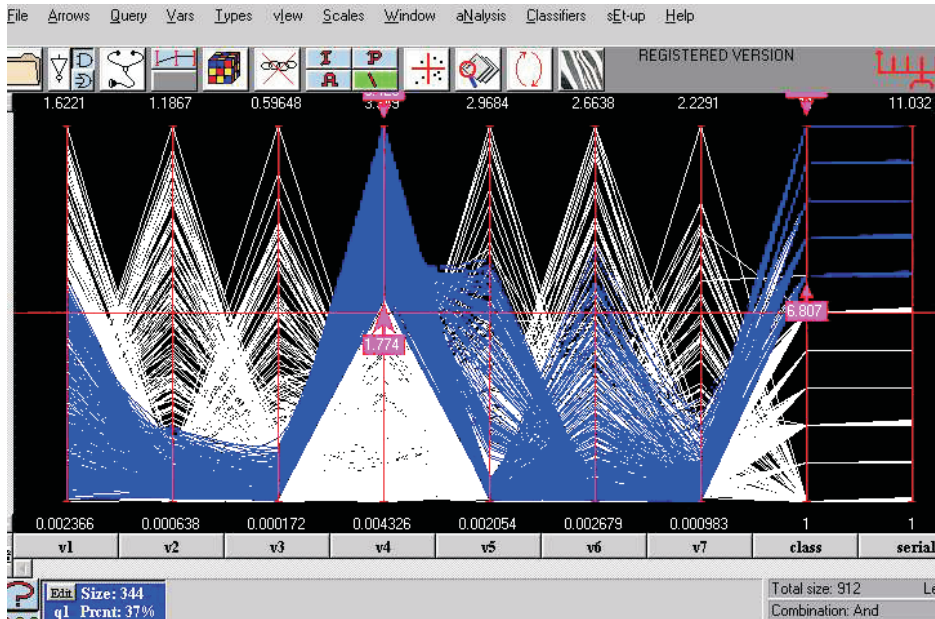


Figure 10.35. Automatic transmissions.

Cluster of vehicles S_1 manufactured by country R having automatic transmissions. This is the classifier's input.

Examining the vehicles' files from the IDs, a surprising separation is found: those with v_4 values in the top (dense range) have — let's call them S_1 (shown in Fig. 10.35) — automatic transmission, and the remainder (denoted by S_2) with sparse v_4 values have manual transmission! Classifying each of the S_1 , S_2 separately and then taking the *union* of the rules provides a much more accurate overall rule for S .

Proceeding, the classification rule for subset S_1 is obtained with an error of 2%; Fig. 10.36. Variable v_6 is redundant and can be omitted. The two variables with the highest predictive value are v_4 , v_2 , and their plot indicates that they are three clusters in the data set: a dense one with low v_2 and v_4 lower than S_1 , and a sparse one with higher values of v_2 . Even though the rule's error is very low, this is a situation in which information on the misclassified trucks is very important. For this purpose, we stop the iteration after the first step. The output is the subset shown in Fig. 10.37. It consists of data from the top five categories, i.e., the subset S and some from category #2. Knowing the classifier's final output is obtained by subsequent iterations refining this subset; we surmise that the misclassifieds are likely to be from class #2. Independently, it was found that the class #2 trucks, manufactured

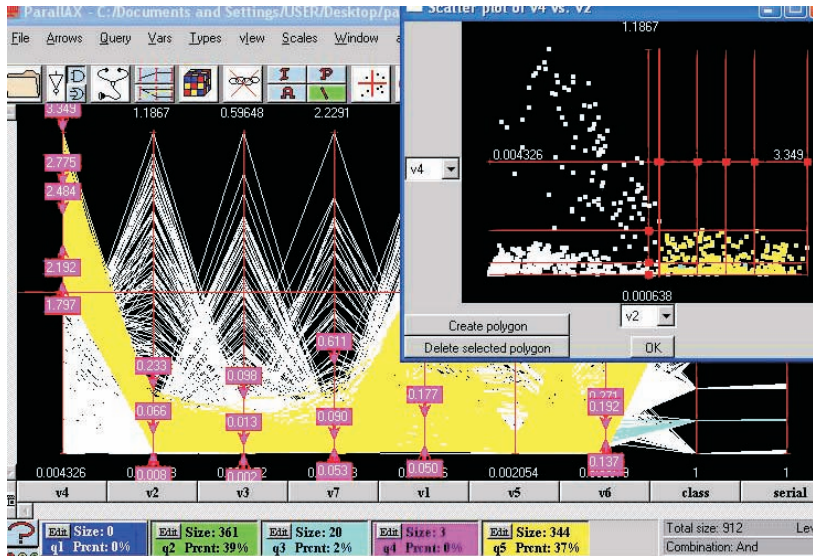


Figure 10.36. Output: the classification rule with a 2% error is found.

Variable v_6 is redundant, and the two best predictors v_4 , v_2 plot shows the three clusters composing the data set.

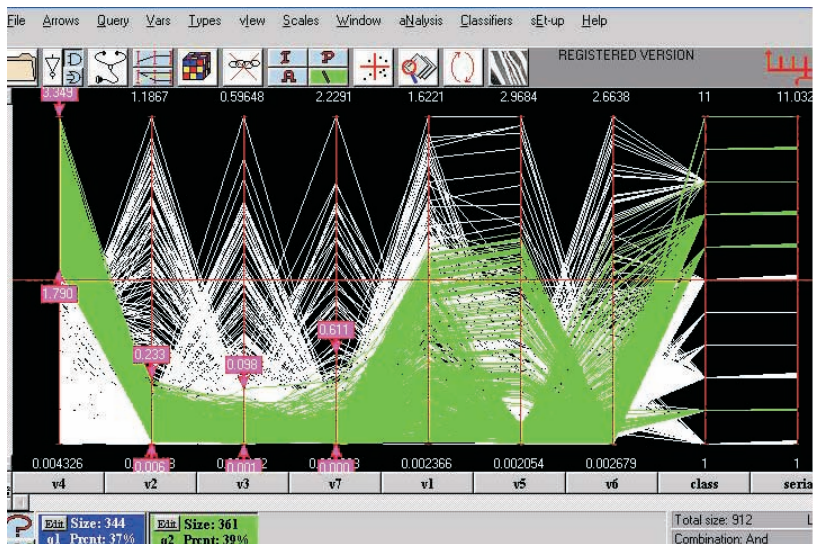


Figure 10.37. The output after the first iteration.

In addition to trucks in the top five classes, the data set contains some from class # 2, so these are the ones likely to be misclassified.

in country A , have been copied by a manufacturer in country R . No wonder, then, that their sound is similar and can be confused with those in S .

10.4 Visual and Computational Models

Finally, we illustrate the methodology's ability to model multivariate relations in terms of hypersurfaces, just as we model a relation between two variables by a planar region. Then by using the interior point algorithm, as shown in Fig. 10.45 of the next section, with the model we can do trade-off analyses, discover sensitivities, understand the impact of constraints, and in some cases do optimization. For this purpose we shall use a data set consisting of the outputs of various economic sectors and other expenditures of a particular (and real) country. It consists of the monetary values over several years for the *agricultural*, *fi*s ing, and *mining* sector outputs, *manufacturing* and *construction* industries, together with *government*, *miscellaneous spending*, and resulting GNP; eight variables altogether. We will not take up the full ramifications of constructing a model from data. Rather, we want to illustrate how $\|\cdot\|$ -coords may be used as a modeling tool. Using the least squares technique, we “fit” a function to this data set and are not concerned at this stage whether the choice is a “good” one. The function obtained bounds a region in \mathbb{R}^8 , 8-dimensional Euclidean space, and is represented by the upper and lower curves shown in Fig. 10.38.

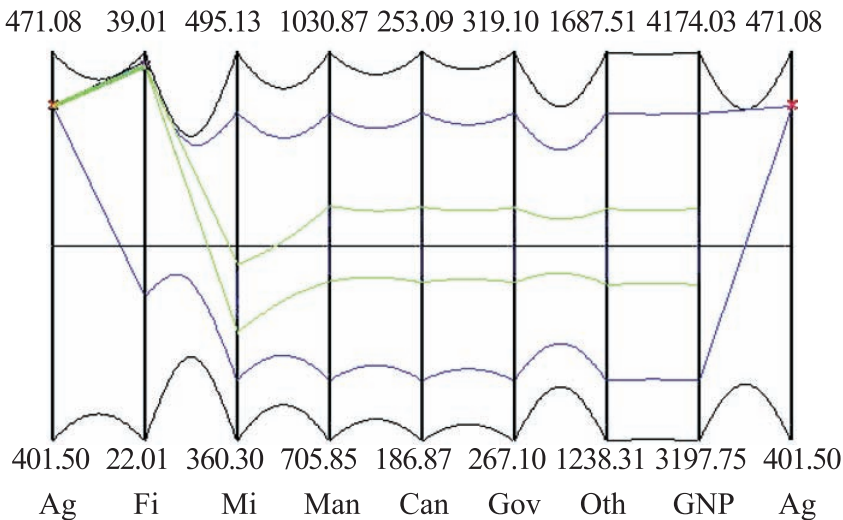


Figure 10.38. Model of a country's economy.

Choosing high *agricultural* and high *fi*s ing output *forces* low *mining* output.

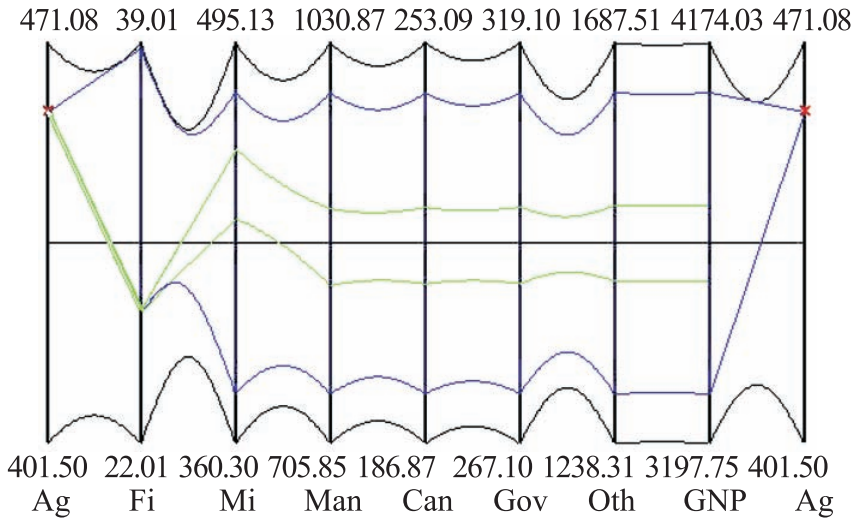


Figure 10.39. Competition for labor between the *fis* ing and *min* ing sectors.

The picture is in effect a simple visual model of the country's economy, incorporating its capabilities, limitations, and interrelationships among the sectors. A point interior to the region satisfies all the constraints simultaneously, and therefore represents a feasible economic policy for that country (i.e., the 8-tuple of values). Using the interior point algorithm we can construct such points. It can be done interactively by sequentially choosing values of the variables, and we see the result of one such choice in Fig. 10.38. Once a value of the first variable is chosen (in this case the *agricultural* output) within its range, the dimensionality of the region is reduced by one. In fact, the upper and lower curves between the second and third axes correspond to the resulting 7-dimensional hypersurface and show the available range of the second variable (*fis* ing) reduced by the constraint. This can be seen (but not shown here) for the rest of the variables. That is, due to the relationship between the eight variables, a constraint on one of them impacts all the remaining ones and restricts their range. The display allows us to experiment and actually see the impact of such decisions downstream. By interactively varying the chosen value for the first variable, we found that it not possible to have a policy that favors *agriculture* without also favoring *fis* ing and vice versa.

Proceeding, a very high value from the available range of *fis* ing is chosen, and it corresponds to very low values of the *min* ing sector. By contrast, in Fig. 10.38 we see that a low value in *fis* ing yields high values for the *min* ing sector. This inverse correlation was examined, and it was found that the country in question has a large number of migrating semiskilled workers. When the fishing industry is doing well,

most of them are attracted to it, leaving few available to work in the mines and vice versa. The comparison between the two figures shows the competition for the same resource between *mining* and *fis* ing. It is especially instructive to discover this interactively. The construction of the interior point proceeds in the same way. In the next section, in the discussion on surfaces this construction is shown for higher-dimensional hypersurfaces.

10.5 Parallel Coordinates: The Bare Essentials

The following short review of \parallel -coords and the discussion on duality provide the essential background to make this chapter self-contained.

10.5.1 Lines

An N -dimensional line ℓ can be described by the $N - 1$ linear equations

$$\ell : \begin{cases} \ell_{1,2} & : x_2 & = & m_2 x_1 + b_2, \\ \ell_{2,3} & : x_3 & = & m_3 x_2 + b_3, \\ & \dots & & \\ \ell_{i-1,i} & : x_i & = & m_i x_{i-1} + b_i, \\ & \dots & & \\ \ell_{N-1,N} & : x_N & = & m_N x_{N-1} + b_N, \end{cases} \quad (10.5)$$

each with a pair of adjacently indexed variables. In the $x_{i-1}x_i$ plane, the relation labeled $\ell_{i-1,i}$, $N = 2, \dots, N$, is a line, and by the *line* \leftrightarrow *point* duality (10.3), it can be represented by the point

$$\bar{\ell}_{i-1,i} = \left(\frac{1}{(1 - m_i)} + (i - 2), \frac{b_i}{(1 - m_i)} \right). \quad (10.6)$$

Here the interaxis distance is 1, so that $i - 2$ is distance between the y (or \bar{X}_1) and \bar{X}_{i-1} axes. Actually, any $N - 1$ independent equations such as

$$\ell_{i,j} : x_i = m_{i,j} x_j + b_{i,j}, \quad (10.7)$$

can equivalently specify the line ℓ , for (10.7) is the projection of ℓ on the $x_i x_j$ 2-D plane, and $N - 1$ such independent projections completely describe ℓ . There is a beautiful and very important relationship illustrated in Fig. 10.40 (left). For a line ℓ in 3-D, the three points $\bar{\ell}_{12}$, $\bar{\ell}_{13}$, $\bar{\ell}_{23}$ are collinear (this line is denoted by \bar{L}), and

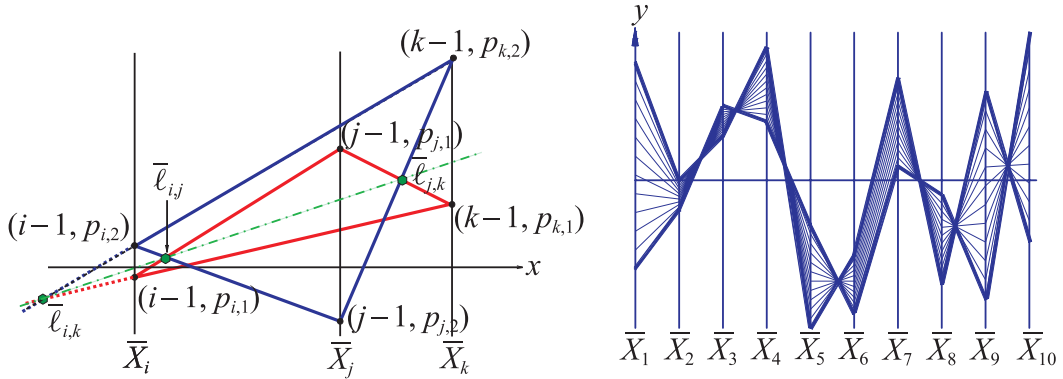


Figure 10.40. Properties of multidimensional lines.

(Left) The three points $\bar{\ell}_{i,j}$, $\bar{\ell}_{j,k}$, $\bar{\ell}_{i,k}$ are collinear for $i \neq j \neq k$. (Right) A line interval in 10-D.

any two points represent ℓ . It is easy to see that a polygonal line on all $N - 1$ points given by (10.6) or their equivalent represents a point on the line ℓ . Conversely, two points determine a line ℓ . Starting with the two polygonal lines representing the points, the $N - 1$ intersections of their \bar{X}_{i-1} , \bar{X}_i portions are the $\bar{\ell}_{i-1,i}$ points for the line ℓ . A line interval in 10-D and several of its points are seen on the (right) of Fig. 10.40. By the way, the indexing of the points $\bar{\ell}$ is essential. It is usually omitted to save display space but it must always be available for it is needed for algorithms that involve lines (e.g., intersections and minimum distance; see Chapter 4).

10.5.2 Planes and Hyperplanes

While a line can be determined from its projections, a plane even in 3-D cannot. A new approach is called for [47]. Rather than discerning a p -dimensional object from its points, it is described in terms of its $(p - 1)$ -dimensional subsets constructed from the points. Let us see how this works. In Fig. 10.41 (left), polygonal lines representing a set of coplanar points in 3-D are seen. From this picture even the most persistent pattern-seeker cannot detect any clues hinting at a relation among the three variables, much less a linear one. The plane has dimension $p = 2$, so we look at *lines* (having dimension $p - 1 = 1$) on the plane constructed so that from each pair of polygonal lines, the lines $\bar{\ell}$ of the 3-point collinearity shown in Fig. 10.40 (left) are obtained. The result, shown on the right, is stunning. All the $\bar{\ell}$ lines intersect at a point, which turns out to be characteristic of coplanarity but not enough to specify the plane. Translating the first axis \bar{X}_1 to the position $\bar{X}_{1'}$, one unit to the right of the \bar{X}_3 axis, and repeating the construction, based on the

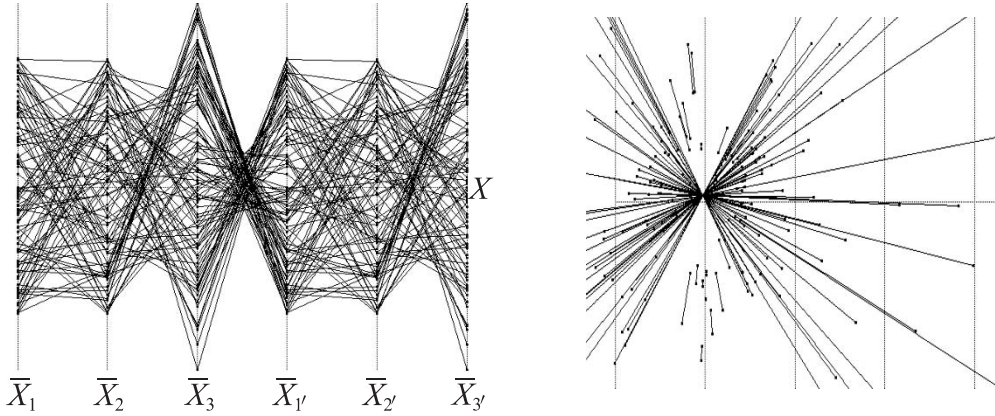


Figure 10.41. Coplanarity.

(Left) The polygonal lines on the first three axes represent a set of coplanar points in 3-D. (Right) Coplanarity! Forming lines on the plane, with the 3-point collinearity, the resulting lines intersect at a point.

axis triple $\bar{X}_2, \bar{X}_3, \bar{X}_{1'}$, yields a second point shown in Fig. 10.42 (left). For a plane described by

$$\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0, \quad (10.8)$$

the two points, in the order they are constructed, are respectively

$$\bar{\pi}_{123} = \left(\frac{c_2 + 2c_3}{S}, \frac{c_0}{S} \right), \quad \bar{\pi}_{1'23} = \left(\frac{3c_1 + c_2 + 2c_3}{S}, \frac{c_0}{S} \right), \quad (10.9)$$

for $S = c_1 + c_2 + c_3$. Three subscripts correspond to the three variables appearing in the plane's equation and the axis triple used for their construction, and distinguish them from the points with two subscripts, representing lines. The second and third axes can also be consecutively translated, as indicated in Fig. 10.41 (left), repeating the construction to generate two more points, denoted by $\bar{\pi}_{1'2'3}, \bar{\pi}_{1'2'3'}$. These points can also be found in an easier way (see Chapter 5). The gist of all this is shown in Fig. 10.42 (right). The distance between successive points is $3c_i$. The equation of the plane π can actually be read from the picture!

In general, a hyperplane in N dimensions is represented uniquely by $N - 1$ points, each with N indices. There is an algorithm that constructs these points *recursively*, raising the dimensionality by one at each step, as is done here starting from points (0-dimensional) and constructing lines (1-dimensional). By the way, all the nice higher-dimensional projective dualities such as *point* \leftrightarrow *hyperplane*

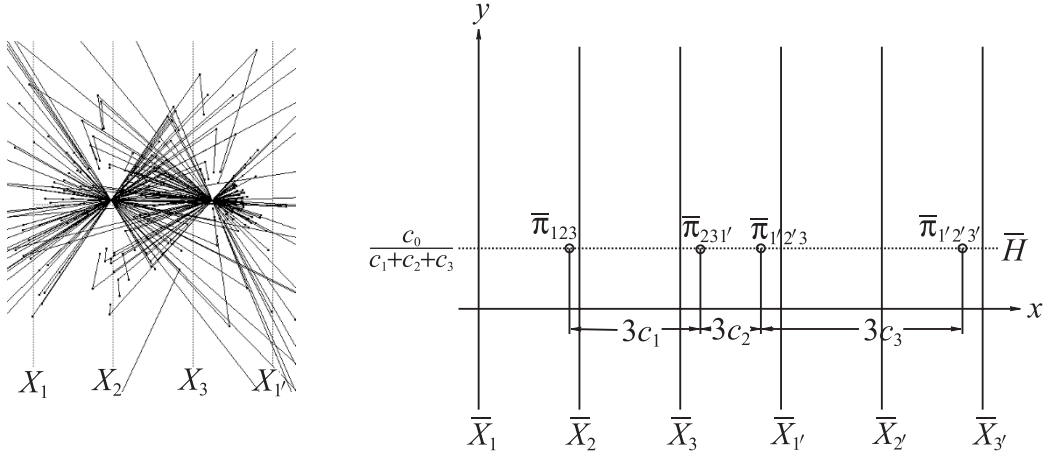


Figure 10.42. Plane representation.

(Left) The two points where the lines intersect uniquely determine a plane π in 3-D. (Right) From four points, similarly constructed by consecutive axis translation, the coefficients of $\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0$, where the coefficients are normalized ($c_1 + c_2 + c_3 = 1$), can be read from the picture!

and *rotation* \leftrightarrow *translation*, hold. Further, a multidimensional object represented in \parallel -coords can still be recognized after it has been acted on by projective transformation (i.e., translation, rotation, scaling, and perspective). The recursive construction and its properties are at the heart of the \parallel -coords method for the visualization of multivariate relations.

Challenge: Visualizing Families of Proximate Planes

Returning to 3-D, it turns out that for points as in Fig. 10.41 that are “nearly” coplanar (i.e., have small errors), the construction produces a pattern very similar to that in Fig. 10.42 (left). A little experiment is in order. Let us return to the family of *proximate* (i.e., close) planes generated by

$$\Pi = \{\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0, \quad c_i \in [c_i^-, c_i^+], \quad i = 0, 1, 2, 3\}, \quad (10.10)$$

randomly choosing values of the c_i within the allowed intervals to determine several planes $\pi \in \Pi$, keeping at first $c_0 = 1$, and plotting the two points $\bar{\pi}_{123}, \bar{\pi}_{1'2'3}$ as shown in Fig. 10.43 (left). Not only is closeness apparent, but more significantly, the distribution of the points is not chaotic. The outline of two hexagonal patterns can be discerned. The family of “close” planes is visualizable, but also the variations in several directions. It is possible to see, estimate, and compare errors or proximity.

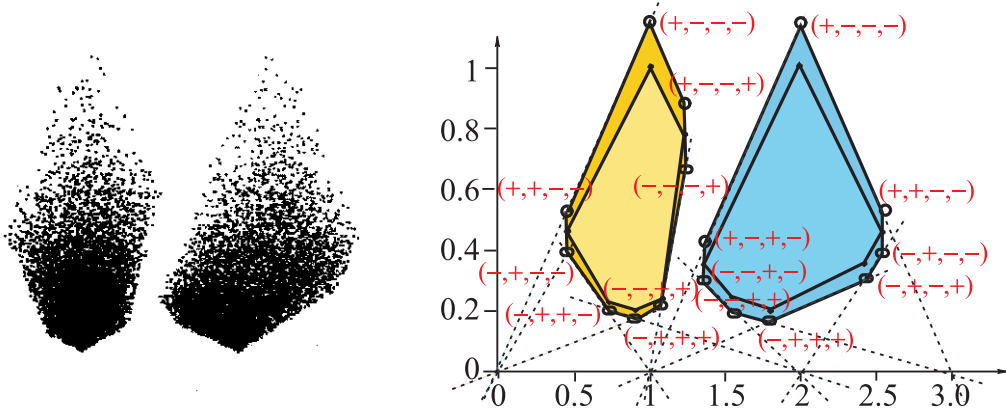


Figure 10.43. A family of close planes.

(Left) A pair of point clusters representing close planes. (Right) The hexagonal regions (interior) are the regions containing the points $\bar{\pi}_{123}$ (left) and $\bar{\pi}_{1'2'3}$ for the family of planes with $c_0 = 1$ and $c_1 \in [1/3, 1.5]$, $c_2 \in [1/3, 2.5]$, $c_3 \in [1/3, 1]$. For c_0 varying, here $c_0 \in [0.85, 1.15]$, the regions (exterior) are octagonal with two vertical edges.

It can be proved that in 3-D, the set of pairs of points representing the family of proximate planes forms two convex hexagons when $c_0 = 1$, with an example shown in Fig. 10.43 (right), and are contained in octagons, each with two vertical edges for varying c_0 . In general, a family of proximate hyperplanes in N -D is represented by $N - 1$ convex $2N$ -gons when $c_0 = 1$, or $2(N + 1)$ -gons for c_0 varying (for an extensive exposition, see Chapter 8). These polygonal regions can be constructed with $O(N)$ computational complexity. Choosing a point in one of the polygonal regions, an algorithm matches the remaining possible $N - 2$ points, one each from the remaining convex polygons, which represent and identify hyperplanes in the family by $N - 1$ points.

We pose the thesis that visualization is not about seeing many things but rather discovering *relations* among them. While the display of randomly sampled *points* from a family of proximate hyperplanes is utterly chaotic (the mess in Fig. 10.41 (right) from points in just *one* plane), their *proximate coplanarity relation* corresponds to a clear and compact pattern. With $\|$ -coords we can focus and *concentrate* the relational information rather than wallowing in the details, ergo the remark “without loss of information” when referring to $\|$ -coords. This is the methodology’s real strength and where the future lies. Here, then, is a visualization challenge: how else can proximate coplanarity be detected and seen?

10.5.3 Nonlinear Multivariate Relations: Hypersurfaces

A relation among two real variables is represented geometrically by a unique region in 2-D. Analogously, a relation between N variables corresponds to a hypersurface in N -D, hence the need to say something about the representation of hypersurfaces in \parallel -coords. A smooth surface in 3-D (and also N -D) can be described as the envelope (discussed in Chapter 6) of all its tangent planes. This is the basis for the representation shown in Fig. 10.44 (left). Every point of the surface is mapped into the two points representing its *tangent plane at the point*. This generates two planar regions, and for N -D there are $N - 1$ such regions. These regions are *linked*, just like the polygons above, to provide the proper $N - 1$ points representing each tangent hyperplane and from which the hypersurface can be reconstructed. Classes of surfaces can be immediately distinguished from their \parallel -coords display (see the chapter on surfaces for extensive treatment). For developable surfaces the regions consist of boundary curves only with no interior points; regions for ruled surfaces have grids consisting of straight lines; quadric surfaces have regions with conic boundaries. These are some examples.

There is a simpler but inexact surface representation that is quite useful when used judiciously. The polygonal lines representing points on the boundary are plotted and their envelope “represents” the surface; the “ ” are a reminder that this is not a *unique* representation. In Fig. 10.45 (left) are the upper and lower envelopes for a sphere in 5-D consisting of four overlapping hyperbolas, which must be distinguished from those in Fig. 10.44 (right), which is exact and interestingly

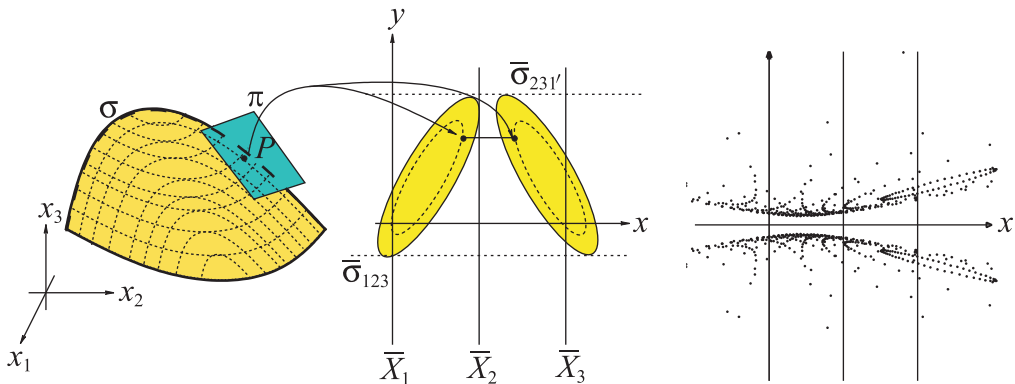


Figure 10.44. Surface representation.

(Left) A smooth surface σ is represented by two planar regions $\bar{\sigma}_{123}$, $\bar{\sigma}_{231}'$ consisting of pairs of points representing its tangent planes. (Right) One of the two hyperbolic regions representing a sphere in 3-D.

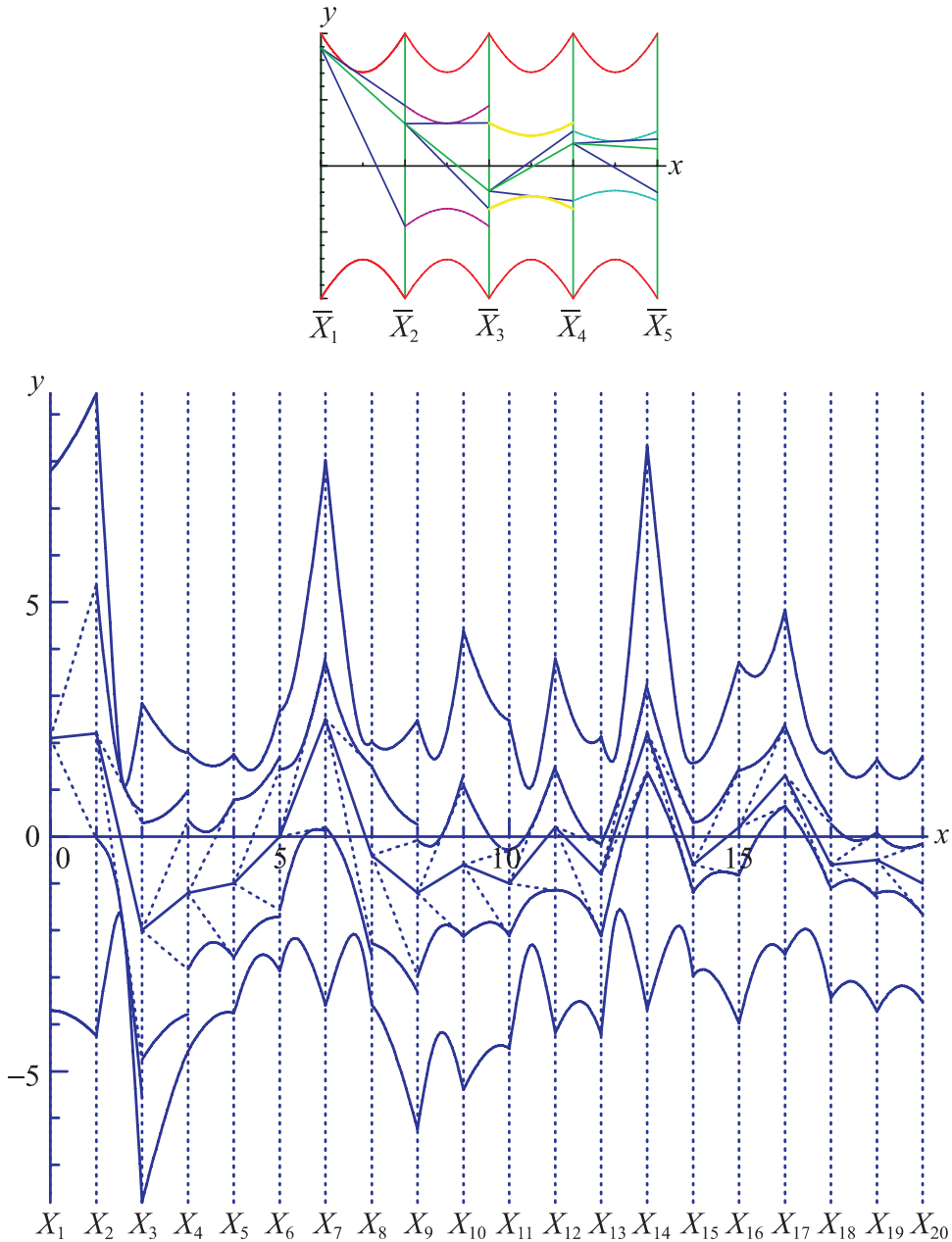


Figure 10.45. Interior point construction.

(Left) A sphere in 5-D showing the construction of an interior point (polygonal line). (Right) The general interior point (polygonal line) construction algorithm shown for a convex hypersurface in 20-D.

enough, are also hyperbolas, the curves determined by points representing the sphere's *tangent planes*. Retaining the exact surface description (i.e., its equation) internally, interior points can be constructed and displayed as shown for the 5-D sphere in Fig. 10.45 (left). On the right the same construction is shown, but for a more complex 20-dimensional convex hypersurface ("model"). The intermediate curves (upper and lower) also provide valuable information and previews of coming attractions. They indicate a neighborhood of the point (represented by the polygonal line) and provide a feel for the local curvature. Note the narrow strips around X13, X14, X15 (as compared to the surrounding ones), indicating that at this state, these are the critical variables where the point is bumping the boundary. A theorem guarantees that a polygonal line that is between all the intermediate curves/envelopes represents an interior point of the hypersurface, and all interior points can be found in this way. If the polygonal line is tangent at any one of the intermediate curves, then it represents a boundary point, while if it crosses any one of the intermediate curves, it represents an exterior point. The latter enables us to see, in an application, the first variable for which the construction failed and what is needed to make corrections. By varying the choice of value over the available range of the variable interactively, sensitive regions (where small changes produce large changes downstream) and other properties of the model can be easily discovered. Once the construction of a point is completed, it is possible to vary the values of each variable and see how this affects the remaining variables. So one can do trade-off analysis in this way and provide a powerful tool for, decision support, process control, and other applications. As new data becomes available the model can be updated, with decisions being made based on the most recent information. This algorithm is used in the earlier example on a model for a country's economy shown in Figs. 10.38, 10.39.

10.6 Future

Searching for *patterns* in a \parallel -coords display is what skillful exploration is about. If there are multivariate relations in the data set, the patterns *are there*, though they may be covered by the overlapping polygonal lines, and that is not all. Our vision is not multidimensional. We do not perceive a room that is 3-dimensional from its points, which are 0-dimensional, but from the 2-dimensional planes that enclose and define it. The recursive construction algorithm does exactly that for the visualization of p -dimensional objects from their $(p - 1)$ -dimensional subsets, one dimension less. We advocate including this algorithm within our armory of interactive analysis tools. Whatever p -dimensional relations exist are revealed by the pattern from the representation of the tangent hyperplanes of the corresponding hypersurface. The

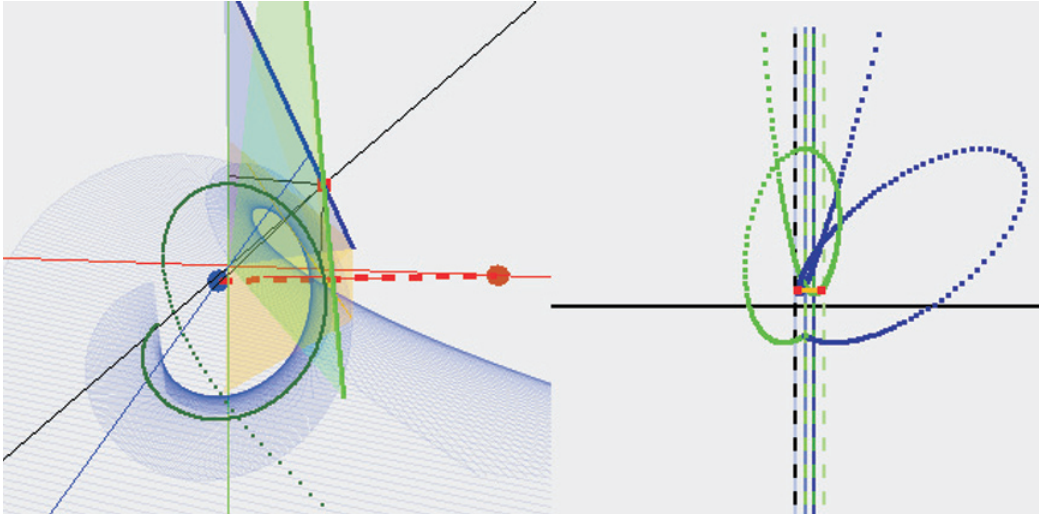


Figure 10.46. A 3-D helicoid in Cartesian and $\|$ -coords.

The two intersecting lines (left) specify one of the helicoid's tangent planes represented by a pair of points one on each of the curves (right). A helicoid in N -D is represented by $N - 1$ such curves.

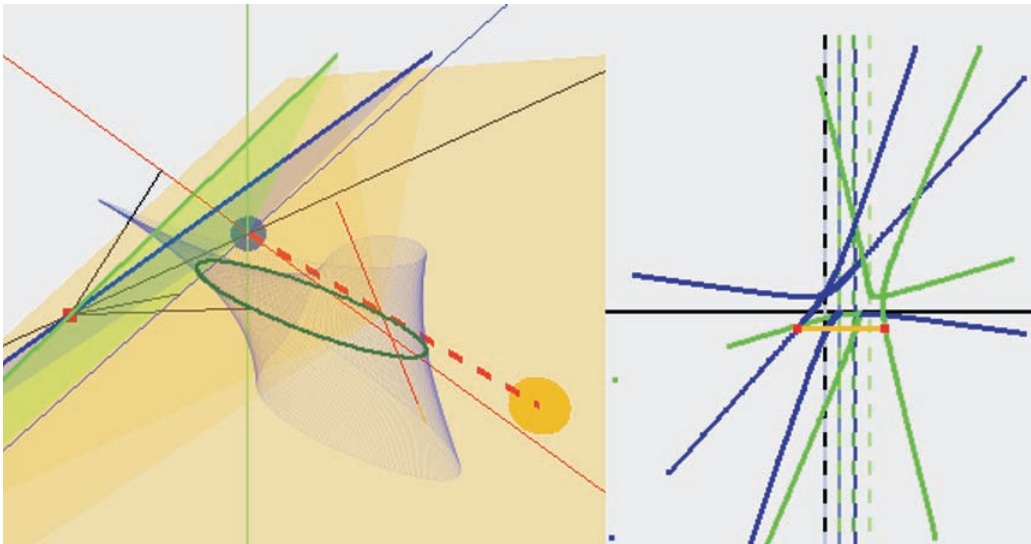


Figure 10.47. A 3-D Möbius strip in Cartesian and $\|$ -coords.

The two intersecting lines (left) specify one of the strip's tangent planes represented by a pair of points, one on each of the curves (right). A path on a Möbius strip in N -D is represented by $N - 1$ such curves. Also see figures in Chapter 9 on Surfaces.

polygonal lines are completely discarded, for the *relation is concentrated in the pattern*: linear relations into points, proximate coplanarity into convex polygons, quadrics into conics, and so on. Note further, again with reference to Figs. 10.40 and 10.41, that relational information resides at the *crossings*. What can be achieved for the representation of complex relations by patterns is exemplified by the pictures of a helicoid and Möbius strip in Cartesian and \parallel -coords in Figs. 10.46, 10.47. These are state-of-the-art results, showing what is achievable and how easily it generalizes to N -D. Can one imagine a higher-dimensional helicoid, much less a *nonorientable* surface like the Möbius strip? It is possible to do such a process on a data set, though at present it is computationally slow. The challenge is to speed up the algorithm for real-time response and thus break the gridlock of multidimensional visualization. There will still be work and fun for the multidimensional detectives visually separating and classifying the *no longer hidden* regions, identifying complex multivariate relations that have evaded us until now.

This page intentionally left blank

Recent Results

11.1 Displaying Several Lines Efficiently

*Shlomi Cohen-Ganor
Department of Computer Sciences
Tel Aviv University
shalmak@yahoo.com*

11.1.1 Abstract

This chapter suggests ways to improve the visualization of several lines in parallel coordinates, specifically how to overcome the inconvenience caused by the fact that the indexed points are not necessarily arranged “neatly” along the horizontal axis. Several approaches are presented by which the input data (or alternatively, the axes) may be transformed so as to reduce or eliminate the problem, with minimal information loss.

11.1.2 Introduction

Recall that in order to visualize a line $\ell \in \mathbb{R}^N$, we first represent it as a set of equations of the following form:⁵⁵

⁵⁵It is also possible to represent the linear relations between the components in different notation, for example,

$$x_i = m_i x_1 + b_i$$

for $i \in [2, \dots, N]$, where x_1 corresponds to some “base” variable. Such representation may convey important properties of the data, but the corresponding graphic representation inherently lacks the “neatness” we are seeking in this chapter.

$$x_i = m_i x_{i-1} + b_i$$

for $i \in [2, \dots, N]$. The line is then visualized in parallel coordinates by $N - 1$ indexed points as follows (in homogeneous coordinates):

$$\bar{\ell}_{i,i+1} : ((i-2)(1-m_i) + 1, b_i, 1-m_i).$$

If ℓ has constant components, they are first plotted independently, using any convenient graphic notation; then the above approach is used for the nonconstant components.

Some basic properties of the indexed points:

- For any pair of points on ℓ , the lines connecting their x_i and x_{i+1} components intersect at $\bar{\ell}_{i,i+1}$.
- The horizontal position of a point $\bar{\ell}_{i,i+1}$ is determined by the corresponding slope m_i .
- The vertical position of a point is determined by the slope as well as the corresponding displacement b_i .

Let us refine the relation between the slope m_i and the horizontal position of $\bar{\ell}_{i,i+1}$:

- If $m_i < 0$, the point lies between the x_i and x_{i+1} axes.
- If $m_i = 0$, the point lies on the x_{i+1} axis.
- If $0 < m_i < 1$, the point lies to the right of the x_{i+1} axis.
- If $m_i = 1$, the point is ideal.
- If $m_i > 1$, the point lies to the left of the x_i axis.

Several problems are apparent:

1. A visualization of ideal points is required for some lines.
2. The indexed point $\bar{\ell}_{i,i+1}$ may fall anywhere in the 2D display, regardless of i or of any other indexed point. Indeed, different indexed points may fall on the same point in 2D. In other words, there is no visual separation between dimension components.
3. If we wish to plot multiple lines in this manner, the display quickly becomes very confusing.
4. There is a discontinuity along the horizontal display axis as the slope approaches 1: A line with slope $1 + \varepsilon$ would appear at the extreme left side, while a line with slope $1 - \varepsilon$ would appear at the extreme right side.

These problems are demonstrated in Figures 11.1 and 11.2. Different ways to overcome these problems are suggested in what follows.

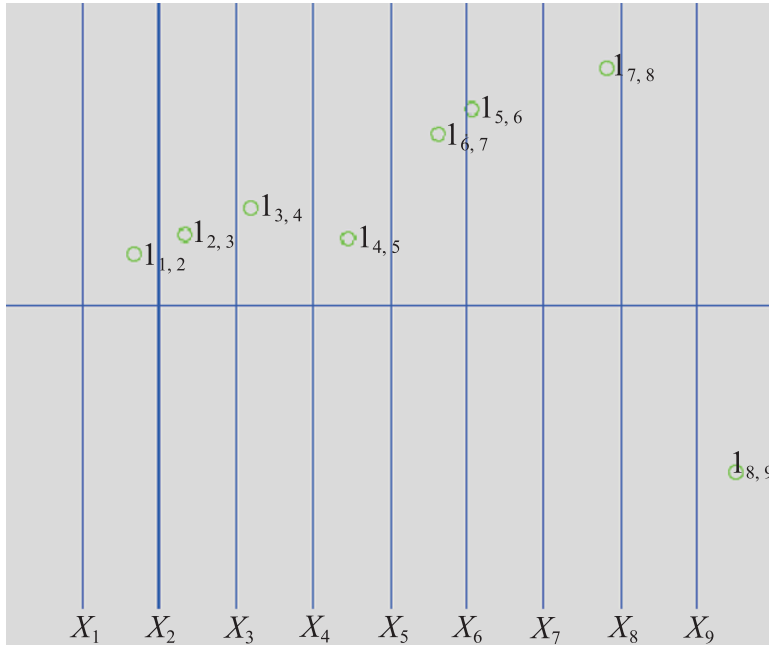


Figure 11.1. A line in \mathbb{R}^9 (9-dimensional Euclidean space).

Note that the first four indexed points are easy to find and understand, while the last four are “mixed up” and confusing. This is because the slopes corresponding to the latter points are positive.

11.1.3 General Approach

Apparently, the easiest way to improve the representation of lines is to invent a new and simpler way of plotting the line data. Specifically, one could easily map the pair (m_i, b_i) onto the area bounded by the axes x_i and x_{i+1} using any continuous and one-to-one function $f : \mathbb{R}^2 \rightarrow (0, 1] \times \mathbb{R}$. This approach would avoid the problems mentioned.

However, an arbitrary choice of the function f would also lose many important properties of the indexed point approach, perhaps most importantly the ability to detect n -flats from a set of points. We would like to construct a mapping that is essentially the same as that described in the previous section, but without having indexed points outside the area bounded by their corresponding axes. As mentioned, this occurs naturally when the slope m_i is negative, indicating that as one variable (dimension) increases, the other decreases and vice versa. Let us seek ways to “make the slope negative.”

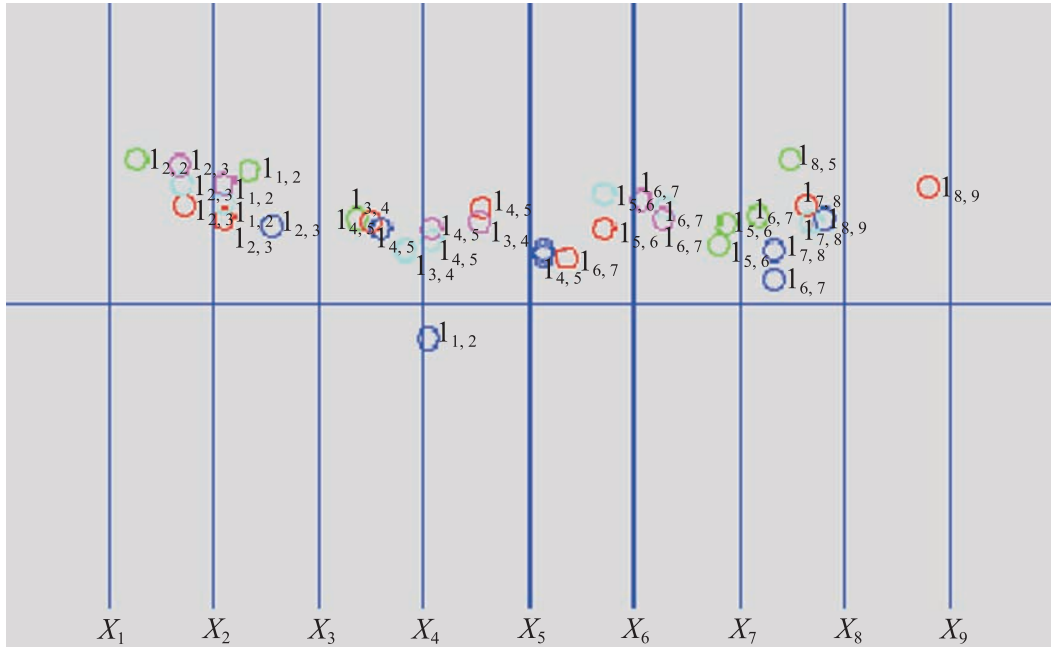


Figure 11.2. Five randomly chosen lines in \mathbb{R}^9 . The display is very hard to understand.

11.1.4 Flipping Axes

As described in an earlier chapter, the up–down directions of the axes are arbitrary, and may be flipped for display purposes, as long as we provide some indication of which vertical direction is positive and which is negative. Flipping an axis means negating the corresponding dimension of the input data, so the slopes from and to that dimension are also negated.

This idea leads to the “axis flipping” algorithm for one line, described in Chapter 4. Let us recall the essential steps:

- For each $i \in [2, N]$:
 - if $m_i > 0$ flip axis i , negate m_i , b_i , and m_{i+1}
 - plot $\bar{\ell}_{i,i+1}$

This works perfectly for a single line, as shown in Figure 11.3. What if we have multiple lines? We could run the algorithm separately for each line, and if we were lucky, the axes would be flipped the same way for all of them, in which case we would have a solution. However, this is not generally the case. Let us investigate the possibilities, first introducing some simple definitions.

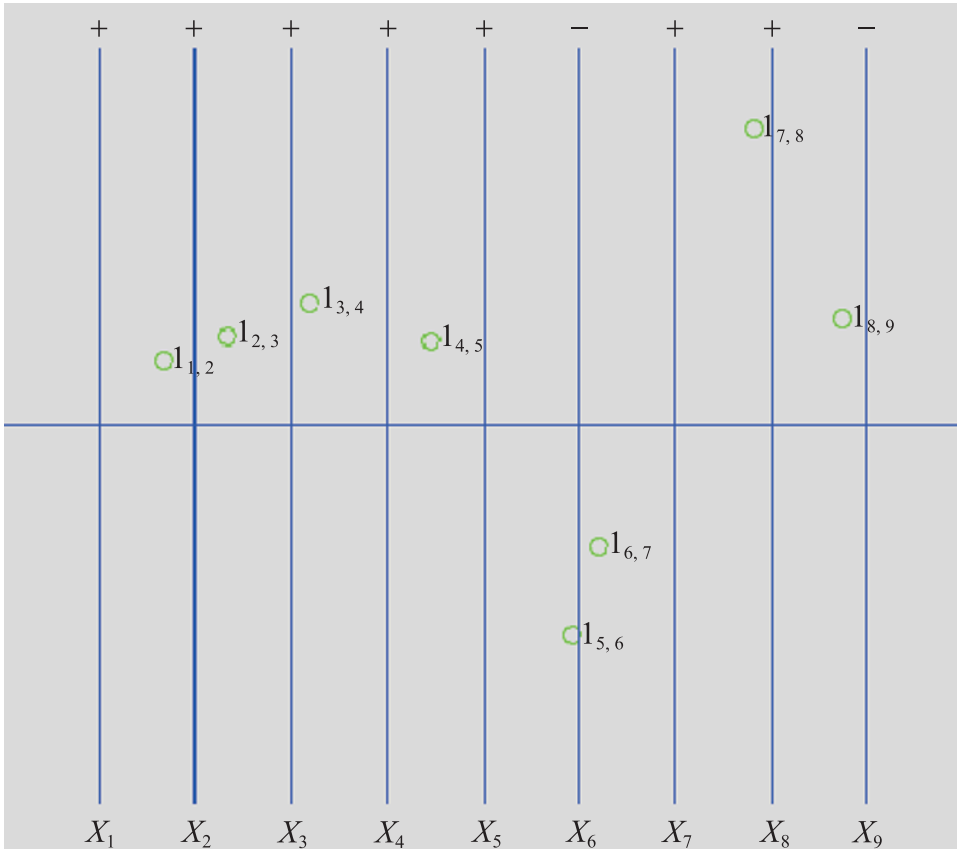


Figure 11.3. Displaying the line shown in Figure 11.1 after flipping axes x_6 and x_9 .

Agreement and Disagreement of Lines and Dimensions

Let us consider a set L of lines in \mathbb{R}^N . For each pair of dimensions $(i, j) \in N \times N$, and for each line $\ell \in L$, we can find the slope m_{ij}^ℓ of the line's projection on the x_i, x_j plane.

Definition We say that lines $\ell, k \in L$ agree on (i, j) if $m_{ij}^\ell m_{ij}^k > 0$ (meaning m_{ij}^ℓ and m_{ij}^k have the same sign).

Intuitively, if each line describes a linear relation between the variables, then the direction of change of j as i increases/decreases is the same in both relations. The agreement relation of lines has the following properties:

1. For any given $(i, j) \in N \times N$, line agreement is an equivalence relation on L .
2. Flipping any axis does not change line agreement (if i or j is flipped, both m_{ij}^l and m_{ij}^k are negated).
3. If lines ℓ, k agree on i, j , then the indexed point $\bar{\ell}_{i,j}$ is between the axes i, j if and only if the indexed point $\bar{k}_{i,j}$ is between the axes i, j .

Recall that we would like all indexed points to be between the axes, and we can achieve this for any one line by flipping axes. It is evident that if all lines in L agree on i, j , we can simultaneously get all the indexed points corresponding to i, j where we want them.

Definition *We say that dimensions i, j agree on L if all lines in L agree on them.*

The agreement relation of dimensions has the following properties:

1. If dimensions i, j agree on L then for any line $\ell \in L$, $\bar{\ell}_{i,j}$ is between the axes i, j if and only if all the indexed points corresponding to i, j are between the axes i, j . This derives directly from property 3 of line agreement.
2. If dimensions i, j do not agree on L , there is at least one line ℓ such that $\bar{\ell}_{i,j}$ is not between the axes i, j (Otherwise, all lines agree on i, j by property 3 of line agreement, and so the dimensions agree by definition).
3. Flipping any axis does not change dimensions agreement (since it does not change lines agreement).

This proves the following conclusion:

- If every pair of adjacent dimensions i, j agree for L , then running the “axis flipping” algorithm on any line in L yields a solution for all lines in L .
- Otherwise, no solution based on flipping exists.

One may wonder whether the order of the axes is important, since the conclusion only mentions adjacent dimensions. Perhaps we could try running the algorithm on different permutations of the axes. However, all of the properties presented above hold for any pair of dimensions and not just adjacent ones. Furthermore, let us redefine agreement of lines as a relation between dimensions:

Definition *We say that dimensions $(i, j) \in N \times N$ agree on lines $\ell, k \in L$ if $m_{ij}^\ell m_{ij}^k > 0$ (meaning m_{ij}^ℓ and m_{ij}^k have the same sign).*

This is an equivalence relation on $N \times N$ for any $\ell, k \in L$. Specifically, it is transitive, since the sign of m_{ij}^ℓ is determined by the signs of m_{id}^ℓ and m_{dj}^ℓ for any dimension d .

It follows that dimension agreement is also transitive and therefore is also an equivalence relation on $N \times N$. So, we can restate our conclusion more generally:

- If all dimensions agree for L , then running the axis flipping algorithm on any line in L , for any permutation of the axes, will give a solution for all lines in L .
- Otherwise, no solution based on flipping or permuting the axes exists.

This suggests a method for finding a partial solution in the general case (not necessarily the best partial solution, but a simple one nonetheless). We can easily find the equivalence classes of the dimensions and reorder the axes by the classes. Running the “axis flipping” algorithm on these axes (on any line) will give a solution such that indexed points corresponding to dimensions from the same class will be between their appropriate axes. This solution maximizes the number of dimensions without any “bad” indexed points, though not necessarily the total number of “bad” indexed points.

To conclude this section, we have shown that the simple “axis flipping” algorithm finds a full solution if one exists, or a partial one otherwise. Reordering the axes may improve the partial solution, but will never uncover a full solution. The basic algorithm is of order $O(N)$, and checking that the solution is correct for all lines is of order $O(L \times N)$, which is the lower bound for the problem. We have also shown a necessary and sufficient criterion for the existence of a solution.

So far we have allowed flipping axes and reordering them. Next we will investigate other operations that can be used.

11.1.5 Rotating Axes

We have proved that the flipping/reordering approach is sufficient if all dimensions *agree* on the set of lines, meaning that for each pair of lines l, k and each pair of dimensions i, j , the slopes m_{ij}^l and m_{ij}^k have the same sign. For brevity we will simply say that the lines *agree* (on any pair of dimensions), which means the same thing by definition.

Let us find a geometric interpretation of this criterion. We are interested only in slopes, so let us translate all lines so that they pass through the origin. Ignore lines parallel to one of the Cartesian axes (see note on constant components in the introduction).

In 2D, agreement of lines means that all lines after translation are contained in exactly two quadrants. More specifically, either they all pass through quadrants I and III, or through quadrants II and IV; see Figure 11.4.

This seems somewhat arbitrary. Consider a group of lines in 2D that have some positive and some negative slopes, but all have a very small absolute slope (as shown

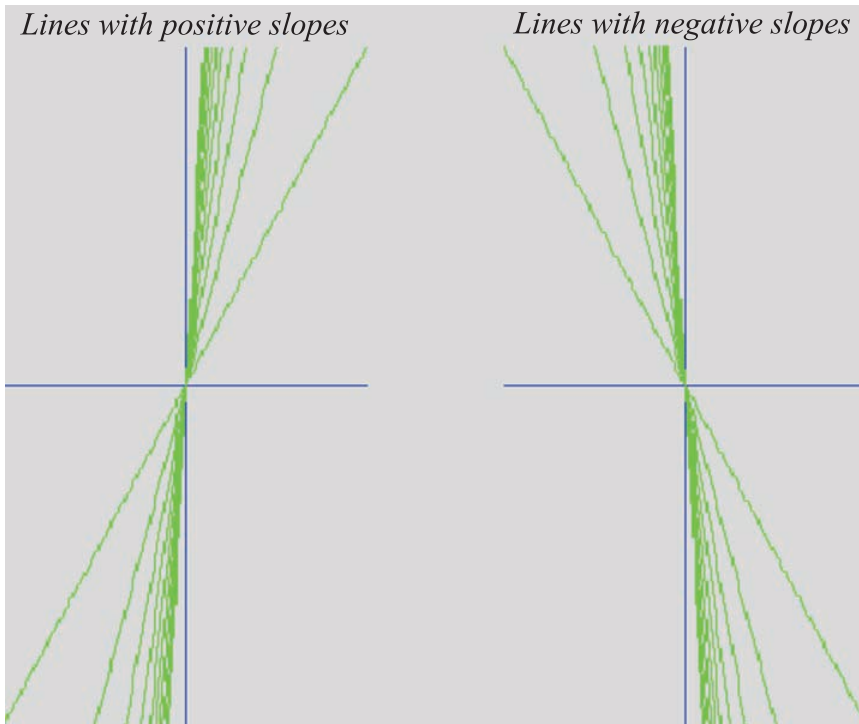


Figure 11.4. Agreeing lines in 2D.

in Figure 11.5). These lines do not *agree*, but they can be made to agree by slightly *rotating the axes*.

More generally, a set of lines in 2D can be made to agree by rotation if and only if they can be bounded between two orthogonal lines. An appropriate rotation angle can be found by taking the angle between one of the bounding lines and one of the axes.

Checking this new criterion is easy, because if a pair of bounding lines exists, then a pair exists such that the slope of one of the lines is just below⁵⁶ one of the given lines. We can look for solutions as follows:

- For each given line, position a bounding line just below it.
- Construct the orthogonal bounding line and count the given lines bounded between them.
- Choose the bounds that contain the most given lines.

⁵⁶“Just below” means “smaller by some infinitesimal value.”

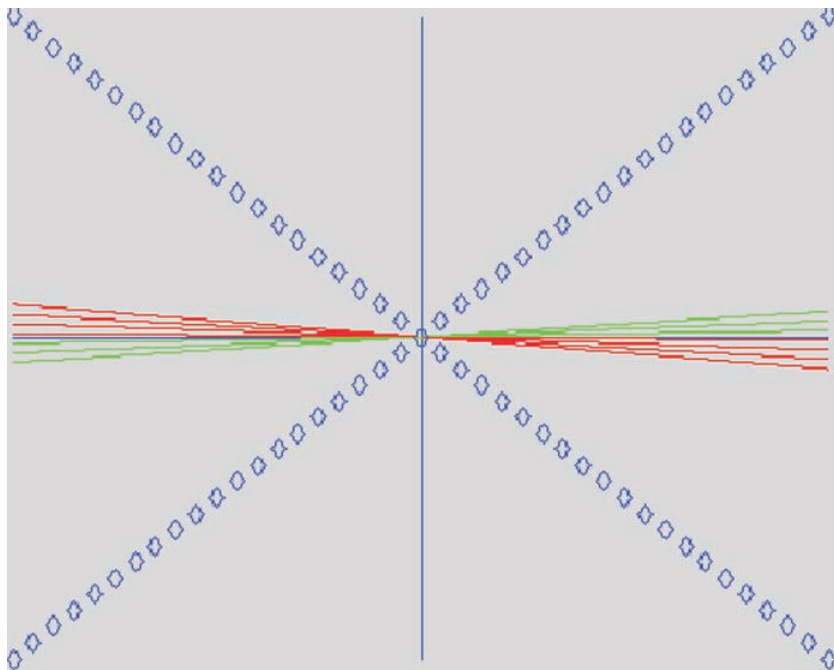


Figure 11.5. Disagreeing lines in 2D.

The green lines have positive slopes, while the red lines have negative slopes. The diamonds show one way to bound all lines between two orthogonal lines.

This method yields a rotation angle such that the greatest number of lines have no “bad” indexed points at all. It also allows us to choose some lines that must not have “bad” indexed points, or to assign priorities to the lines, etc.

It is worth noting that orthogonal lines never agree, and they cannot be made to agree by any rotation (since they will always remain orthogonal).

All that was said is true for 2D. In 3D, the agreement criterion means that all lines are contained in exactly two octants. We would like to find a 3D rotation, if one exists, such that a given set of lines will be contained in exactly two octants after the rotation.

As in 2D, a set of lines in 3D can be made to agree by rotation if and only if they can be bounded between three orthogonal lines (this is not an exact definition). In this case, an appropriate rotation matrix can be constructed by inverting the matrix that maps the Cartesian axes onto unit vectors on the bounding lines.

Unlike the 2D case, it is not clear how one can find the bounding lines, or check whether they exist. This question is left open.

The subject of rotations in 3D and higher dimensions will be revisited later on.

11.1.6 Transforming Axes

We have seen evidence that applying a rotation to the data is useful in some cases for achieving a better visualization. While this may not be acceptable in some applications, rotated data preserves many properties of the original data. Let us now take a step further in this direction and allow a broader class of transformations. Since we are interested in representing lines, we will require that the transformation preserve lines. We do not want to lose any information, so we will also require that the transformation be invertible. The class of invertible linear transformations seems a natural direction to investigate.

Let us represent lines as vectors: for each given line, construct a vector from the origin in the same direction as the line. For convenience and without loss of generality, we will construct the vector so that the first nonzero component is positive. A set of m lines in \mathbb{R}^N can now be represented by an $m \times n$ matrix V , where each column holds the vector representation of one line. The transformation is represented by an $n \times n$ matrix T that is nonsingular (this guarantees an invertible transformation).

For the lines to agree, the matrix TV should have the same sign for all components in the same row. Without loss of generality, we will require the matrix TV to have only positive cells. Geometrically, this is equivalent to requiring all lines to pass through two specific quadrants/octants etc., which is more than we need, but the notation becomes simpler and no genuine solutions are lost.

Now we must find a nonsingular T such that $TV > 0$ (cell by cell). Equivalently, $V^T T^T > 0$ (cell by cell). To find T^T we must find N linearly independent vectors $x \in \mathbb{R}^N$ satisfying $V^T x > 0$ (cell by cell). These vectors are the rows of T . We will now describe an algorithm for finding such vectors.

Recall that the first column of V^T is nonnegative because we took vectors whose first nonzero component is positive. For now, we assume that the first column is positive:

- Assign $x_1 = e_1$
- Choose any $\varepsilon > 0$
- For $j \in [2, \dots, N]$ do:
 - Find $t_j = \max \left\{ \frac{\varepsilon - V_{ij}^T}{V_{i1}^T} : 1 \leq i \leq m \right\} \cup \{0\}$
 - Assign $x_j = t_j * e_1 + e_j$

Proof of correctness:

The assumption guarantees that $V^T x_1 > 0$ (cell by cell).

From the construction,

$$(V^T x_j)_i = \sum_{k=1}^N V_{ik}^T x_{kj} = t_j V_{i1}^T + V_{ij}^T \geq \frac{\varepsilon - V_{ij}^T}{V_{i1}^T} V_{i1}^T + V_{ij}^T = \varepsilon > 0.$$

It is clear that the matrix T^T is upper triangular with determinant 1, so T is nonsingular.

If the assumption on the first column does not hold, i.e., there are zeros, it is always possible to find a rotation matrix R such that $(RV)^T$ satisfies the assumption (this can be shown by constructing an infinitesimal rotation such that the zeros become nonzeros, but the other cells remain nonzeros). The rotation can then be considered part of the linear transformation. Alternatively, if some other column does not contain zeros, it can be made the first column by reordering the dimensions, and it can be made positive by reversing some of the unit vectors as necessary.

We have found a solution for any given set of lines! In conclusion, if invertible linear transformations of the data are allowed, we can transform any set of lines into agreeing lines, which can then be perfectly visualized by the “axis flipping”

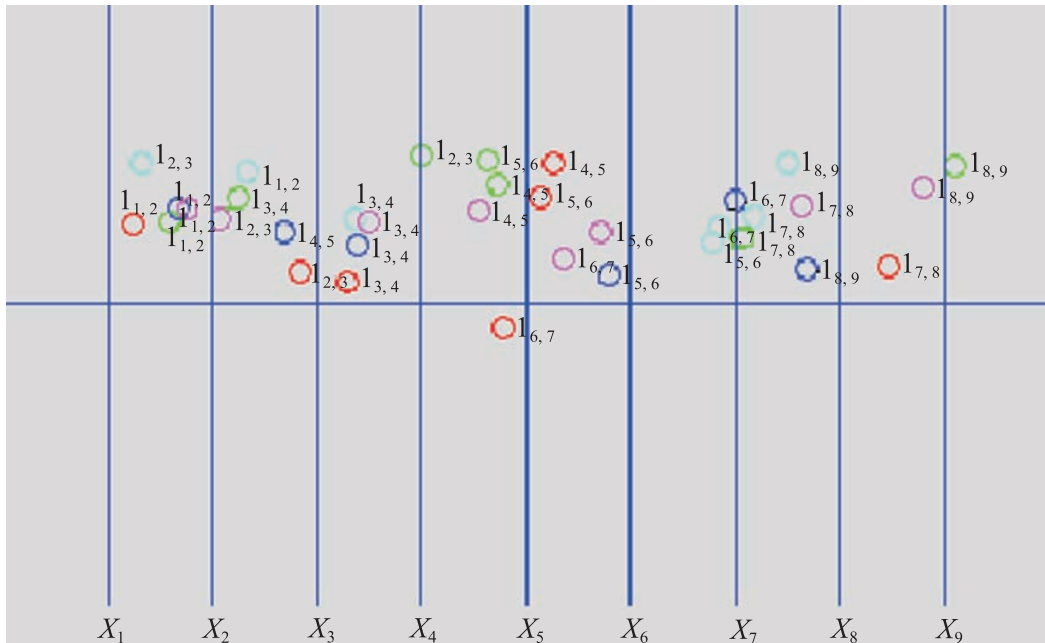


Figure 11.6. Five randomly chosen lines in \mathbb{R}^9 (one ideal point is not displayed).

algorithm. Actually, we could modify the matrix so that no flipping is required at all, since this is just a special case of invertible linear transformations.

The method will be exemplified using five random lines in \mathbb{R}^9 , shown in Figure 11.6.

The corresponding vector matrix is

$$V = \begin{pmatrix} .51 & .19 & .34 & .49 & .12 \\ -.38 & -.62 & -.18 & .12 & -.04 \\ -.19 & .14 & -.22 & .29 & .51 \\ -.44 & -.34 & .37 & -.52 & -.53 \\ .16 & -.07 & .65 & -.57 & .43 \\ .59 & .44 & -.18 & -.25 & -.17 \\ .05 & .8 & 0.0 & .04 & -.44 \\ -.92 & .26 & .07 & -.17 & .25 \\ -.09 & .22 & .29 & -.52 & -.06 \end{pmatrix}.$$

Applying the construction algorithm with $\varepsilon = 0.5$ yields

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5.8947 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.1176 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 8.5833 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 3.0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 5.5833 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 7.8333 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 2.7843 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 4.6667 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

And the transformed vectors are

$$V = \begin{pmatrix} .51 & .19 & .34 & .49 & .12 \\ 2.6263 & 0.5 & 1.8242 & 3.0084 & 0.6674 \\ 0.89 & 0.5424 & 0.5 & 1.3276 & 0.7641 \\ 3.9375 & 1.2908 & 3.2883 & 3.6858 & 0.5 \\ 1.69 & 0.5 & 1.67 & 0.9 & 0.79 \\ 3.4375 & 1.5008 & 1.7183 & 2.4858 & 0.5 \\ 4.045 & 2.2883 & 2.6633 & 3.8783 & 0.5 \\ 0.5 & 0.789 & 1.0167 & 1.1943 & 0.5841 \\ 2.29 & 1.1067 & 1.8767 & 1.7667 & 0.5 \end{pmatrix}.$$

Indeed, all cells are positive, as expected. Notice that the first row remains unchanged, and that every other row contains ε at least once. A simple way to

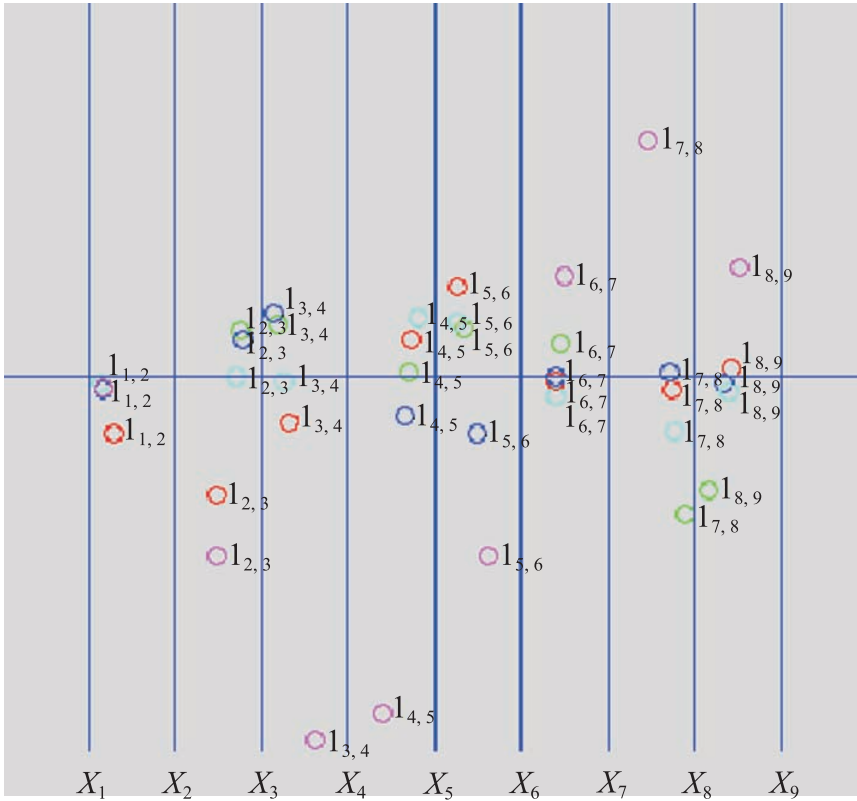


Figure 11.7. The same five lines after transformation. The visual display is much clearer than before.

get the transformed lines is to choose a pair of points on each line, apply the transformation T to the points, and construct the lines passing through the points correspondingly. The result can be seen in Figure 11.7. The visual display is much clearer than before.

We can do even better by noticing that the indexes are now redundant: all points between axes X_i, X_{i+1} have indexes $i, i + 1$. The final result is shown in Figure 11.8.

11.1.7 Back to Rotating Axes

Now that we have found a general solution, we would still like to find a solution based only on rotation and flipping, since this preserves more properties of the data (particularly angles).

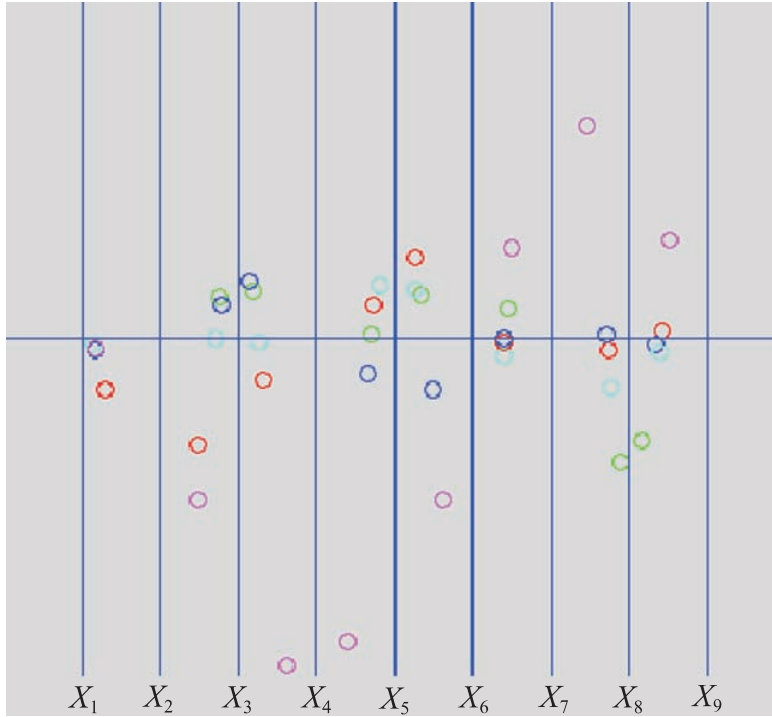


Figure 11.8. The same five lines after transformation, without the indexes.

In terms of linear algebra, we still want to find N linearly independent vectors $x \in \mathbb{R}^N$ satisfying $V^T x > 0$ cell by cell, but we also want them to be orthogonal.⁵⁷ This would guarantee that the resulting transformation matrix T is not only invertible but also orthogonal, which is the same as saying that it is a rotation matrix. Geometrically, the criterion $V^T x > 0$ defines an open polyhedron. We know that the hyperplanes defining it all pass through the origin, so the origin must be an extreme point. We also know that the polyhedron is unbounded; multiplying a point inside the polyhedron by any positive number yields another point inside the polyhedron. We are seeking a set of N points inside the polyhedron such that the lines connecting them with the origin are orthogonal. Such a set exists if and only if it is possible to place an N -cube so that one of the cube's vertices is at the origin and all other vertices are inside the polyhedron. This problem is left open.

⁵⁷More precisely, we want them to be orthonormal, but the additional constraint does not affect the problem significantly.

11.2 Separating Point Clusters on Different Planes

Nir Shahaf
School of Mathematical Sciences
Tel Aviv University, Israel
shahafni@post.tau.ac.il

11.2.1 Overview

We start by describing a family of problems. Suppose that we have m planes in \mathbb{R}^N and that we have a sample of n random points taken from each of those planes. Our purpose is to be able to recover the planes from the sample data (without knowing which sample point came from which plane). This family of problems depends on two parameters: m , the number of planes, and N , the dimension. We solve the problem in a certain setting where $m = 3$ and $N = 3$ (extension to other problems from the family is immediate). We show a probabilistic algorithm that solves the problem with high probability; the complexity of that algorithm is $O(n^2)$.

11.2.2 The Algorithm

- All possible pairs of the input points are taken, and for each pair (except for the special case that the polygonal lines representing the two points are parallel), the line joining the two points is found (represented by two points).
- The input for the next step is the lines found in the previous one. The lines are randomly divided into pairs; this is the only random element in the algorithm. For each pair, the intersection point is taken. If the four points from which the two lines originated came from the same plane, the found point will be the point representing the intersection line between the plane and the first superplane. We refer to those points as “superlines”.
- In this step we look for clusters of points. All the “superlines” that were created by points coming from a mutual plane will consolidate on the point representing the intersection line between the plane and the first superplane. At the beginning, our workspace is a portion of the plane containing all the “superlines.” At each step of the computation, the workspace is divided into 100 equally sized portions. For each portion, the number of “superlines” in it is counted. For each portion containing a number of points that is larger than some threshold (which is determined in a way described later), we recursively look for consolidating points in it (taking the portion as the new workspace). Once the total number of points in the workspace does not change for a certain number of recursion

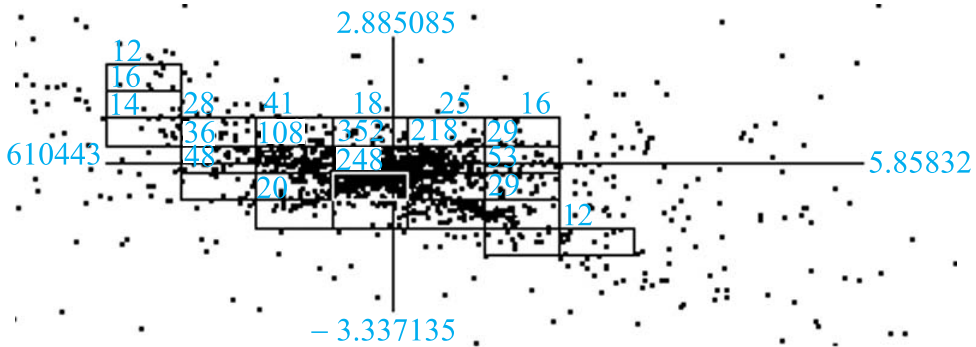


Figure 11.9. Sample run of the algorithm. “Interesting” rectangles are shown.

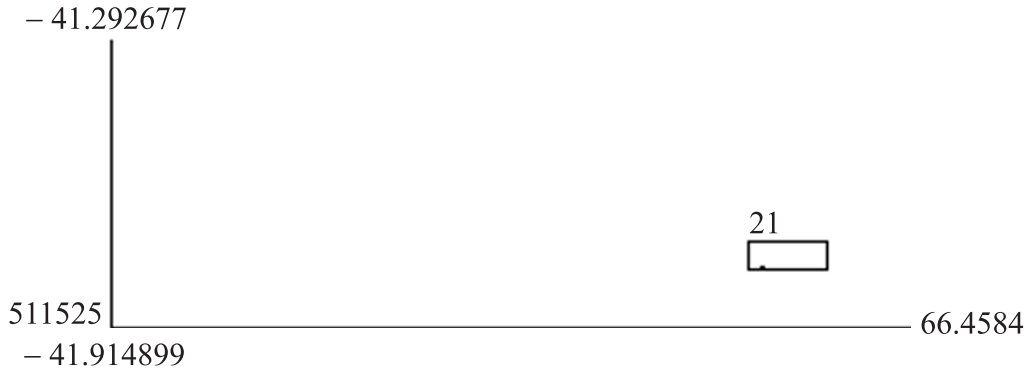


Figure 11.10. A cluster of 21 points is found.

steps (depending on the precision of the calculations), we declare that a point has been found. In this way, three points are found, one for each plane.

The process is performed four times, once for each of the coordinate systems: (X_1, X_2, X_3) , (X_2, X_3, X'_1) , (X_3, X'_1, X'_2) , and eventually (X'_1, X'_2, X'_3) .

- After the previous steps we have the four intersection points of the planes with the four first superplanes; after sorting the point triplets, the planes' equations can be easily found.

One question remains open: how do we determine the best threshold? For obvious efficiency reasons, we would like to choose a threshold that is as high as possible. However, choosing too high a threshold might cause us to miss some of

the intersection clusters. The next subsection deals with the question of determining a threshold that is not too high and not too low.

11.2.3 Mean and Variance Estimation

Let n be the number of random points taken on each of the three planes. The total number of points is $3n$. In the first phase of the computation we have taken all possible pairs; each pair yielded a line. Denote the number of lines by α :

$$\alpha = \binom{3n}{2} = \frac{3n(3n-1)}{2}.$$

Suppose that π is one of the three planes. Since we have taken n points from each plane, the number of lines on the plane π (out of the lines created by taking all pairs of points) would be

$$\beta = \binom{n}{2} = \frac{n(n-1)}{2}.$$

In the second phase of the computation we have paired all the α lines and created $\frac{\alpha}{2}$ “superlines.” Let $Y_1, Y_2, \dots, Y_\alpha$ be random variables; Y_i indicates the contribution of the i th line to the total number of superlines consolidating on the point representing the plane π . If the i th line was not on π , then $Y_i \equiv 0$ (since this line is not on π , in any possible pairing, a superline that was created by this line would not be the point representing π). We can therefore observe only the subset of those random variables corresponding to the β lines on the plane π . Indicate them by X_1, X_2, \dots, X_β . Observing the i th line, this line will contribute $\frac{1}{2}$ to the number of consolidating superlines (its pair would contribute the other $\frac{1}{2}$, and together they contribute their mutual superline) if and only if its pair is one of the β lines as well. After choosing the first element of the pair, there are $\alpha - 1$ elements left in all, and $\beta - 1$ “good” elements left. Therefore $P(X_i = \frac{1}{2}) = \frac{\beta-1}{\alpha-1}$. Denote that probability by p :

$$X_i = \begin{cases} \frac{1}{2} & \text{with probability } p, \\ 0 & \text{with probability } 1 - p. \end{cases}$$

Let X be a random variable indicating the total number of superlines consolidating on the point representing the plane π . Then $X = \sum_{i=1}^{\beta} X_i$. Since $EX_i = \frac{p}{2}$, we have that

$$EX = E \sum_{i=1}^{\beta} X_i = \sum_{i=1}^{\beta} EX_i = \sum_{i=1}^{\beta} \frac{p}{2} = \frac{p\beta}{2}.$$

In order to calculate the variance of X , we use the fact that

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{\beta} X_i\right) = \sum_{i=1}^{\beta} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq \beta} \text{cov}(X_i, X_j).$$

Since $2X_i$ is Bernoulli distributed, we have that

$$4\text{Var}(X_i) = \text{Var}(2X_i) = p \cdot (1 - p) = \frac{\beta - 1}{\alpha - 1} \cdot \frac{\alpha - \beta}{\alpha - 1},$$

$$\text{Var}(X_i) = \frac{(\alpha - \beta)(\beta - 1)}{4(\alpha - 1)^2}.$$

For $i \neq j$ we have that $\text{cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$, and therefore we need to find the distribution of $X_i X_j$. Since both X_i and X_j have only two possible values (0 and $\frac{1}{2}$), $X_i X_j$ has the same property and may have only the following two values, 0 and $\frac{1}{4}$:

$$\begin{aligned} P\left(X_i X_j = \frac{1}{4}\right) &= P\left(X_i = \frac{1}{2}, X_j = \frac{1}{2}\right) \\ &= P\left(X_i = \frac{1}{2}\right) \cdot P\left(X_j = \frac{1}{2} \mid X_i = \frac{1}{2}\right). \end{aligned}$$

We already know that $P(X_i = \frac{1}{2}) = p$. We need to evaluate $P(X_j = \frac{1}{2} \mid X_i = \frac{1}{2})$. Assuming that the i th line was paired with a “good” line, then the j th line will be paired with a “good” line as well if and only if one of the following occurs: the j th line was paired with the i th line, or it was paired with a different “good” line. The probability for the first event is $\frac{1}{\alpha-1}$ (we want to choose a specific element out of the $\alpha - 1$ available elements left). The probability for the second is $\frac{\beta-3}{\alpha-3}$ (we have already excluded three elements: the i th line, its pair, and the j th line). Therefore

$$\begin{aligned} P\left(X_i X_j = \frac{1}{4}\right) &= p \cdot \left(\frac{1}{\alpha - 1} + \frac{\beta - 3}{\alpha - 3}\right) \\ &= \frac{\beta - 1}{\alpha - 1} \cdot \frac{(\alpha - 3) + (\beta - 3)(\alpha - 1)}{(\alpha - 1)(\alpha - 3)} \\ &= \frac{\beta - 1}{\alpha - 1} \cdot \frac{\alpha - 3 + \alpha\beta - \beta - 3\alpha + 3}{(\alpha - 1)(\alpha - 3)} \\ &= \frac{(\beta - 1)(\alpha\beta - \beta - 2\alpha)}{(\alpha - 1)^2(\alpha - 3)}, \end{aligned}$$

and therefore

$$E(X_i X_j) = \frac{(\beta - 1)(\alpha\beta - \beta - 2\alpha)}{4 \cdot (\alpha - 1)^2(\alpha - 3)}.$$

Obviously,

$$E(X_i) \cdot E(X_j) = \frac{p^2}{2^2} = \frac{(\beta - 1)^2}{4 \cdot (\alpha - 1)^2}.$$

Hence

$$\begin{aligned} \text{cov}(X_i, X_j) &= E(X_i X_j) - E(X_i) \cdot E(X_j) \\ &= \frac{(\beta - 1)(\alpha\beta - \beta - 2\alpha)}{4 \cdot (\alpha - 1)^2(\alpha - 3)} - \frac{(\beta - 1)^2}{4 \cdot (\alpha - 1)^2} \\ &= \frac{\alpha\beta^2 - \beta^2 - 2\alpha\beta - \alpha\beta + \beta + 2\alpha - (\beta - 1)^2(\alpha - 3)}{4 \cdot (\alpha - 1)^2(\alpha - 3)} \\ &= \frac{\alpha\beta^2 - \beta^2 - 2\alpha\beta - \alpha\beta + \beta + 2\alpha - (\beta^2 - 2\beta + 1)(\alpha - 3)}{4 \cdot (\alpha - 1)^2(\alpha - 3)} \\ &= \frac{\alpha\beta^2 - \beta^2 - 2\alpha\beta - \alpha\beta + \beta + 2\alpha - \alpha\beta^2 + 3\beta^2 + 2\alpha\beta - 6\beta - \alpha + 3}{4 \cdot (\alpha - 1)^2(\alpha - 3)} \\ &= \frac{2\beta^2 - \alpha\beta - 5\beta + \alpha + 3}{4 \cdot (\alpha - 1)^2(\alpha - 3)}, \end{aligned}$$

and eventually,

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^{\beta} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq \beta} \text{cov}(X_i, X_j) \\ &= \frac{\beta(\alpha - \beta)(\beta - 1)}{4(\alpha - 1)^2} + 2 \cdot \frac{\beta(\beta - 1)}{2} \cdot \frac{2\beta^2 - \alpha\beta - 5\beta + \alpha + 3}{4 \cdot (\alpha - 1)^2(\alpha - 3)}. \end{aligned}$$

Corollary 11.2.1. *Let X be a random variable indicating the number of superlines consolidating on the point representing the intersection between the plane π and a superplane. Then*

$$EX = \frac{p\beta}{2}, \quad (11.1)$$

$$\text{Var}(X) = \frac{\beta(\alpha - \beta)(\beta - 1)}{4(\alpha - 1)^2} + 2 \cdot \frac{\beta(\beta - 1)}{2} \cdot \frac{2\beta^2 - \alpha\beta - 5\beta + \alpha + 3}{4 \cdot (\alpha - 1)^2(\alpha - 3)}. \quad (11.2)$$

11.2.4 Error Probability Approximation

In the second phase of the computation, the pairing phase, we pair the α lines to obtain $\frac{\alpha}{2}$ “superlines.” Let X be the random variable defined in the previous subsection. We are now interested in the following question: what is the probability that X is exactly k ?

A different approach can be taken in order to describe the pairing of the α lines. Let γ be the number of all possible pairs of lines taken from the α lines:

$$\gamma = \binom{\alpha}{2} = \frac{\alpha(\alpha - 1)}{2}.$$

Let δ be the number of all possible “good” pairs, where both of the lines come from the “good” β elements:

$$\delta = \binom{\beta}{2} = \frac{\beta(\beta - 1)}{2}.$$

Consider the following process: for $\frac{\alpha}{2}$ rounds, we pick pairs from the γ possible pairs. Our choice might be impossible, since two distinct pairs we choose might have a line in common (that would not be a legal pairing). However, every possible pairing can be described as a choice of its pairs out of all the possible pairs. Let Y be a random variable indicating the number of “good” pairs chosen (out of the δ “good” pairs). We indicate the sample space by Ω and define the following events:

$$A = \{\omega \in \Omega \mid \omega \text{ is a legal pairing}\},$$

$$B = \{\omega \in \Omega \mid X(\omega) = k\},$$

$$C = \{\omega \in \Omega \mid Y(\omega) = k\}.$$

Notice that when $X = k$, our choice is a legal pairing, and in addition, we have chosen exactly k “good” pairs: $B = A \cap C$. Therefore: $B \subsetneq C$.

Corollary 11.2.2. *For X and Y define as above,*

$$P(X = k) < P(Y = k).$$

In order to obtain a bound on the probability that $X = k$, we simply need to estimate the probability that $Y = k$. The process we have described is actually taking $\frac{\alpha}{2}$ samples from a population of γ elements, of which δ are considered “good.” Then Y is exactly k if when choosing the $\frac{\alpha}{2}$ elements, we choose exactly k “good” ones and $\frac{\alpha}{2} - k$ “bad” ones (hypergeometric distribution):

$$P(Y = k) = \frac{\binom{\delta}{k} \binom{\gamma - \delta}{\frac{\alpha}{2} - k}}{\binom{\gamma}{\frac{\alpha}{2}}}.$$

If we take our threshold to be $EX - 4\sigma$, where $\sigma = \sqrt{\text{Var}(X)}$, then for two sample cases (where $n = 25$ and $n = 100$), the bound probabilities are listed in Table 11.1 (see also Figures 11.11 and 11.12).

Considering the fact that those probabilities are just bounds, and that the real ones are much lower (by eliminating all the illegal pairings), the method would

Table 11.1. Error probability bounds.

Points Per Plane	EX	$\sqrt{\text{Var}(X)}$	$EX - 4\sigma$	$P(Y < EX - 4\sigma)$
25	16.16798	2.55499	5.94799	0.0011748
100	273.11144	10.47581	231.20816	0.004782

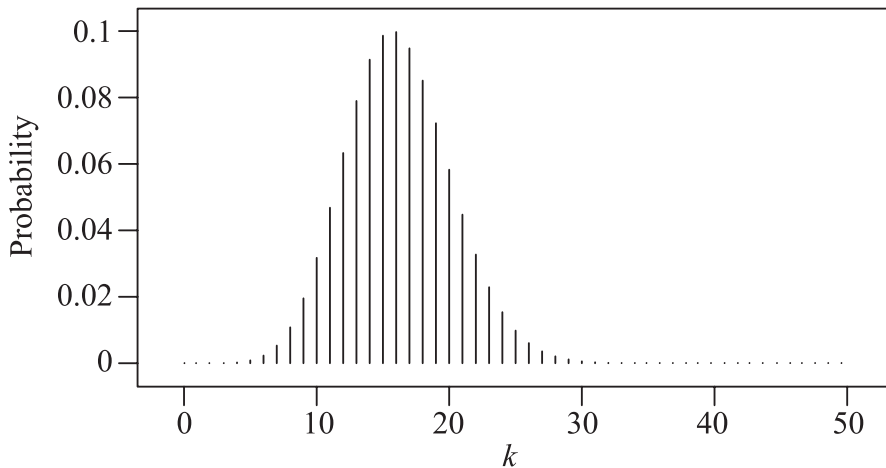


Figure 11.11. $P(Y = k)$, 25 points per plane.

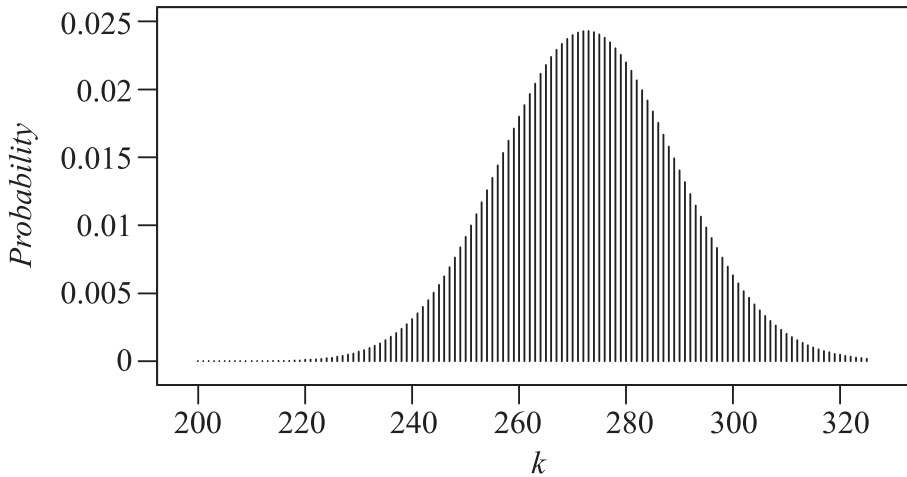


Figure 11.12. $P(Y = k)$, 100 points per plane.

almost never miss the points. Even if it does miss, executing the process once more will succeed with even higher probability.

11.2.5 Complexity

In this subsection we will estimate the complexity of the process described in the previous subsections. In the described algorithm, we first take all possible pairs of the n points. We have $\alpha(n)$ of those, which is of order $O(n^2)$. At the next stage we randomly divide the $\alpha(n)$ lines into pairs, find their intersection point, and get $l(n) := \frac{\alpha(n)}{2}$ “superlines.” Obviously, the complexity of that stage is of order $O(n^2)$ as well. The final stage, the recursive consolidation points lookup, requires a more delicate analysis. Denote by $M(n)$ the following constant:

$$M(n) = \frac{l(n)}{t(n)},$$

where $t(n)$ is the threshold chosen as described in the previous subsections ($EX - 4\sigma$). This constant is a bound on the maximal number of distinct portions of the workspace that can contain a number of points that is more than the threshold at the same time. However, for our purposes, we need a bound that does not depend on n . Fortunately, this is the case. A short computation shows that the sequence $M(n)$ is strictly decreasing; hence there exists a constant M such that for all n ,

$$M(n) < M.$$

Another interesting point (which will not be used here) is that sequence actually converges:

$$\lim_{n \rightarrow \infty} M(n) = 3^4,$$

where the number 3 in the limit is due to the number of planes we started with. During the recursion process, we take care of portions of the original workspace that get smaller: first we start with 10^2 of them; on the second recursion step we already have 10^4 of them, and so on. However, on every recursion level, there cannot be more than $M(n)$ portions that contain more than $t(n)$ points. We now define one more constant, C , which is dictated by the precision of the computation. The constant C is the maximal depth of the recursion that still improves the computation. This is a universal constant, which does not depend on n (it depends only on the capabilities of the computer performing the calculations). Since on every level we can inspect at most $M(n)$ portions of the workspace, and since the maximal recursion level is C , we have that the maximal number of totally inspected portions is bounded by

$$C \cdot M(n) < C \cdot M.$$

Note that this is a very loose bound: it is obviously impossible to be simultaneously in $M(n)$ portions in the C th level. The meaning of that scenario would be finding $M(n)$ points, while we expect to find only 3. While processing a portion we count a number of points that is less than $O(n^2)$ (we observe only the points that belong to the portion). Therefore the complexity of the consolidation points lookup is of order $O(n^2)$, and this is also the complexity of the entire algorithm, as claimed.

11.3 Surface Representation Decomposition and Developable Quadrics

Chao-Kuei Hung
Department of Information Management
Chaoyang University of Technology
Taichung, Taiwan
ckhung@cyut.edu.tw

In the parallel coordinate system, a smooth hypersurface σ in \mathbb{R}^N is identified as the envelope of its tangent hyperplanes. Each of σ 's tangent hyperplanes π gives an indexed set of points $(\bar{\pi}_{12...N}, \bar{\pi}_{1'2...N}, \dots, \bar{\pi}_{1'2'...N})$ in the projective plane \mathbb{P}^2 . The hypersurface σ is thus represented as an indexed set of regions $(\bar{\sigma}_{12...N}, \bar{\sigma}_{1'2...N}, \dots, \bar{\sigma}_{1'2'...N})$ composed of these points.

In this paper, we show that this representation is the composition of an extended version of projective duality and a linear transformation. For a quadric, the computations can be stated explicitly. On the other hand, as pointed out in Chapter 9, each representing region reduces to a curve if the hypersurface is developable. For a hypersurface that is both developable *and* quadric, the decomposition theorem gives rise to an algorithm for computing the equations of the representing curves.

11.3.1 The Hyperplane Representation as a Linear Transformation on the Coefficients

Consider a hyperplane $\pi \in \mathbb{R}^N$ as embedded in \mathbb{P}^N with the equation $c_0x_0 + c_1x_1 + \dots + c_Nx_N = 0$ written in projective coordinates. We will focus on just one of its representing points $\bar{\pi}_{12\dots N} \in P^2$, knowing that all other representing points can be analyzed similarly.

Let $\mathbf{c} = [c_0, c_1, \dots, c_N] \in P^N$, $\mathbf{u} = [1, 1, \dots, 1] \in P^N$, and $\mathbf{d} = [d_1, d_2, \dots, d_N] \in P^N$. We recall (9.5) that $\bar{\pi}_{12\dots N} = [\mathbf{c} \cdot \mathbf{u}, \mathbf{c} \cdot \mathbf{d}, -1]$. Here the positions of the coordinates are slightly rearranged, so that the variable corresponding to the inhomogeneous constant in the Euclidean space is placed first. Now let us define the linear transformation

$$\mathcal{I} : \mathbf{v} \mapsto M_{\mathcal{I}}\mathbf{v},$$

where the matrix $M_{\mathcal{I}}$ is directly constructed from the above equations:

$$M_{\mathcal{I}} = \begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 0 & d_1 & d_2 & \dots & d_N \\ -1 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The $\|\cdot\|$ -coords representation of a hyperplane can thus be seen as an indexed set of such linear transformations applied to the coefficients of the hyperplane. When we want to emphasize that a different index gives rise to a different matrix and thus a different linear transformation, we will index the map by, for example, $\mathcal{I}_{123\dots N}$, indicating that the $d_1, d_2, d_3, \dots, d_N$ entries in the matrix $M_{\mathcal{I}_{123\dots N}}$ are replaced with appropriate alternatives.

11.3.2 A Brief Review of Duality in Projective Space

In the proper terms of projective geometry, a point in the projective space is technically a 1-dimensional vector subspace of \mathbb{R}^{N+1} , while a line is technically a 1-dimensional collection of such vector subspaces, and so on. We will be slightly

sloppy here and refer to a point as a 0-flat, a line as a 1-flat, and so on, as if they were embedded in \mathbb{R}^N .

Let us first recall the well-known “point \longleftrightarrow hyperplane” duality in projective geometry:

$$[c_0, c_1, \dots, c_N] \longleftrightarrow \{[x_0, x_1, \dots, x_N] : c_0x_0 + c_1x_1 + \dots + c_Nx_N = 0\}.$$

More generally, an m -flat dualizes to an $(N - m - 1)$ -flat. We will not make direct use of this general version of duality, and therefore a precise formula is not of our concern. However, we will apply the idea of this generalization to an analogous situation.

Now let us move on to one of the simplest classes of nonflat geometric objects: the quadrics, or quadratic hypersurfaces. The equation of a quadric σ is of the form $\mathbf{x}^T Q \mathbf{x} = 0$, where $\mathbf{x} \in \mathbb{P}^N$ and Q is an $(N + 1) \times (N + 1)$ real symmetric matrix. For $N = 2$, a nondegenerate quadric is a conic section and may be either an ellipse, a parabola, or a hyperbola. Ellipsoids, hyperboloids, cones, and cylinders are all examples of quadrics with $N = 3$.

If Q is nonsingular, then $\mathbf{x}^T Q^{-1} \mathbf{x} = 0$ also defines a quadric $\bar{\sigma}_Q$. In fact, points on Q dualize to tangent hyperplanes of $\bar{\sigma}_Q$, and vice versa. Thus we can understand duality between quadrics σ and $\bar{\sigma}_Q = \sigma_{Q^{-1}}$ in terms of the duality between points and hyperplanes for any σ with nonsingular Q .

These are all familiar results from projective geometry up to this point.

11.3.3 Extending the Projective Duality to Degenerate Quadrics

What if Q is singular? What is a reasonable definition of $\bar{\sigma}_Q$, the dual geometric object of σ in this case? Take the following surface in \mathbb{P}^3 as an example:

$$\sigma : [x_0 \ x_1 \ x_2 \ x_3] \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0.$$

This is the elliptical cylinder $4x_1^2 + 9x_2^2 = 1$ in an inhomogeneous equation. In the same spirit as the nondegenerate case, we will find its tangent hyperplanes, and dualize these to become points that collectively form the “dual geometric object” we seek.

Let $\xi = [\xi_0, \xi_1, \xi_2, \xi_3] \in \sigma$. The hyperplane π_ξ passing through ξ and tangent to σ is

$$-(x_0 - \xi_0)\xi_0 + 4(x_1 - \xi_1)\xi_1 + 9(x_2 - \xi_2)\xi_2 = 0,$$

which reduces to

$$-\xi_0 x_0 + 4\xi_1 x_1 + 9\xi_2 x_2 = 0,$$

since $\xi \in \sigma$. Its dual $\bar{\xi}$ is therefore $[-\xi_0, 4\xi_1, 9\xi_2, 0]$. It is easily verified that the collection of all such points lies on the intersection of the plane $x_3 = 0$ and the quadric $-x_0^2 + \frac{x_1^2}{4} + \frac{x_2^2}{9} = 0$. It is also straightforward to verify that conversely, every point on the intersection dualizes to a tangent plane of σ . The dual of this particular elliptical cylinder is therefore an ellipse (a curve) in \mathbb{P}^3 . Let us call this curve $\bar{\sigma}$.

Are we justified in calling this curve the “dual” of σ ? If we “dualize” this ellipse, do we get the elliptical cylinder back? For a brief moment it is easier to think in \mathbb{R}^3 rather than in \mathbb{P}^3 . Let $\eta = (\eta_1, \eta_2, 0)$ be a point on $\bar{\sigma}$, where we write in inhomogeneous coordinates and η_0 is taken as 1. There are many tangent planes to $\bar{\sigma}$ at η . In fact, they are precisely the pencil of planes containing the tangent straight line to $\bar{\sigma}$ at η . The equation of any plane in this family looks like $\frac{\eta_1}{4}(x_1 - \eta_1) + \frac{\eta_2}{9}(x_2 - \eta_2) + tx_3 = 0$, where $t \in \mathbb{R}$. Indeed, these planes collectively dualize to one ruling of σ .

The above example can be generalized in two ways.

First let us extend this to higher dimensions. Suppose that Q is an $(N + 1) \times (N + 1)$ diagonal matrix of rank $c + 1$ representing a hypersurface σ in \mathbb{P}^N . Without loss of generality, we will assume that the nonzero entries are in the upper left $(c + 1) \times (c + 1)$ block, and we will write this block as $Q_{(c+1)(c+1)}$. That is, $Q = \left(\begin{array}{c|c} Q_{(c+1)(c+1)} & 0 \\ \hline 0 & 0 \end{array} \right)$. The hypersurface σ can be seen as composed of “rulings” each of which is an $(N - c)$ -flat of the form $\{[\xi_0, \xi_1, \dots, \xi_c, x_{c+1}, \dots, x_N] : x_{c+1}, x_{c+2}, \dots, x_N \in \mathbb{R}\}$. All points on one “ruling” share the same tangent hyperplane. Moreover, it is a result from projective geometry that the dual of any hyperplane of this form lies in the c -flat ϕ that is the intersection of the $N - c$ hyperplanes $x_{c+1} = 0, x_{c+2} = 0, \dots, x_N = 0$. (Think of the plane $x_3 = 0$ in the third example.)

We might as well restrict our attention on σ to the confine of ϕ , a c -dimensional subspace of \mathbb{P}^N , since this “cross section” of σ gives all nontrivial features of σ . The rest is just “linear extension” along the $N - c$ uninteresting coordinate axes. Denote $[x_0 \ x_1 \ \dots \ x_c]^T$ by \mathbf{x}_ϕ . The equation of σ restricted to ϕ is thus $\mathbf{x}_\phi^T R \mathbf{x}_\phi = 0$. Since $Q_{(c+1)(c+1)}$ is nonsingular, the dual of σ (in this restriction space) is $\mathbf{x}_\phi^T Q_{(c+1)(c+1)}^{-1} \mathbf{x}_\phi = 0$. Stepping back into \mathbb{P}^N , the dual of σ is seen as a $(c - 1)$ -dimensional object embedded in a c -dimensional subspace of \mathbb{P}^N .

The second generalization invokes results from linear algebra. Let Q be a real symmetric matrix of rank $c + 1$. It is unitarily diagonalizable as $Q = U \Lambda U^T$ for

some orthogonal matrix U and some diagonal matrix Λ . In fact, column vectors of U are the unit eigenvectors of Q , and the diagonal elements of Λ are the corresponding eigenvalues. Thus the equation of an arbitrary quadric $\sigma : \mathbf{x}^T Q \mathbf{x} = 0$ can be written as $\mathbf{x}^T U \Lambda U^T \mathbf{x} = 0$. Perform a change of basis $\mathbf{y} = U^T \mathbf{x}$ to obtain $\mathbf{y}^T \Lambda \mathbf{y} = 0$, and the previous results can then be applied to σ in this new coordinate system, since Λ is diagonal. We note that both tangency and duality of flats are invariant under change of basis. The $c + 1$ column vectors of U corresponding to the nonzero eigenvalues span ϕ , the intersection c -flat. The remaining $N - c$ column vectors correspond to the directions that form the “rulings.”

The result in this section can be summarized as follows. Let σ be a geometric object in \mathbb{P}^N that

- lies in the intersection of e hyperplanes, which is a copy of \mathbb{R}^{N-e} that we shall call ϕ ,
- is composed of “rulings” each of which is a copy of \mathbb{P}^f , where $\bar{\mathcal{L}} \equiv e$,
- has an equation of the form $\mathbf{x}^T Q \mathbf{x} = 0$ in ϕ , where the rank of Q is $c + 2 \equiv N + 1 - e - f$.

Then the duality map \mathcal{D} maps σ to a geometric object $\bar{\sigma}$ in \mathbb{P}^N that

- lies in the intersection of $\bar{e} \equiv f$ hyperplanes, which is a copy of \mathbb{R}^{N-f} that we shall call ψ ,
- is composed of “rulings” each of which is a copy of $\mathbb{P}^{\bar{f}}$, where $\bar{f} \equiv e$,
- has an equation of the form $\mathbf{x}^T \bar{Q} \mathbf{x} = 0$ in ψ , where the rank of \bar{Q} is again $c + 2 = N + 1 - \bar{e} - \bar{f}$.

This is actually a slightly generalized version of what we have discussed above. However, we will make use only of the results from the prior discussions and hence will omit the tedious but straightforward proof of this more general result. On the other hand, this general result is so interesting that such digressions and some further comments are warranted.

We choose the mnemonics e , f , and c to mean, respectively,

- emptiness: the difference of dimensions between N and that of the minimal subspace that contains σ ,
- flatness: the “degree of degeneracy” of σ , or more precisely, the dimension of each flat “ruling” of σ ,
- curvedness: the number of nonvanishing terms in the diagonalized first fundamental form, properly generalized to objects in higher dimensions.

In view of the identity $e + f + c = N - 1$ and the above observation, one can say that *emptiness dualizes to flatness and vice versa, while curvedness is unchanged*

by the duality \mathcal{D} . Consider again the above elliptical cylinder $4x_1^2 + 9x_2^2 = x_0^2$ as an example. Here $e = 0$, $f = 1$, $c = 1$, and $N = 3$. For its dual, a circle in \mathbb{P}^3 , $\bar{e} = 1$, $\bar{f} = 0$, and $\bar{c} = 1$.

As a special case, $c = 0$ means $f + e = N - 1$. This corresponds to the classical duality that a k -flat in \mathbb{P}^N dualizes to an $(N - k - 1)$ -flat. As another special case, $c = N - 1$ means $f = e = 0$. This corresponds to the classical duality for nondegenerate quadrics, whose dual is found by inverting its representing symmetric matrix. The flat duality and the quadric duality in projective geometry are thus unified in this general duality map \mathcal{D} .

We will make use of only the special case $e = 0$ in later sections, and will be particularly interested in the most special case $e = 0$ and $c = 1$, the class of developable hypersurfaces.

11.3.4 The Decomposition and Its Application to Developable Quadrics

The decomposition theorem can now be easily stated.

Theorem 11.3.1. *The parallel-coordinate representation of hypersurfaces using tangent hyperplanes is equivalent to an indexed set of compositions: $\{\mathcal{I}_I \circ \mathcal{D} : I \in \{123 \cdots N, 1'23 \cdots N, 1'2'3 \cdots N, \dots, 1'2'3' \cdots N\}$.*

This is a direct result of the fact that in parallel-coordinate representations, we take tangent hyperplanes to objects and then draw *their* representations in the \parallel -coords plane to represent the original hypersurface. Again for convenience, in the following discussions we will focus on just one of the compositions and omit the indices.

This representation typically gives rise to a patch in the \parallel -coords plane, since typically a smooth hypersurface in \mathbb{P}^N has an $(N - 1)$ -parameter family of tangent hyperplanes. In the special case of developable hypersurfaces, it has a 1-parameter family of tangent hyperplanes, which means that \mathcal{D} maps the hypersurface to a curve in \mathbb{P}^N . Subsequently, \mathcal{I} maps the space curve onto a plane curve in \mathbb{P}^2 . For a developable quadric, this plane curve can be explicitly computed.

Let σ be a developable quadric in P^N defined by the homogeneous equation $\mathbf{x}^T Q \mathbf{x} = 0$, where Q is an $(N + 1) \times (N + 1)$ real symmetric matrix. Its flatness is $N - 2$, and Q has rank 3.

First, unitarily diagonalize Q as $U \Lambda U^T$, where U 's column vectors are the unit eigenvectors of Q and the diagonal elements of Λ are the corresponding eigenvalues. For convenience, we will assume that the only three nonvanishing eigenvalues are arranged to appear at the upper left corner. As a notational convenience, for any matrix A we will

- denote by A_{*3} the submatrix composed of the leftmost three columns of A ,
- denote by A_{3*} the submatrix composed of the topmost three rows of A ,
- and denote by A_{33} the upper left 3×3 submatrix of A .

Thus Λ_{33} is the only nonvanishing part of Λ , and U_{*3} is the submatrix composed of the corresponding eigenvectors.

Perform a change of coordinate system $\mathbf{y} = U^T \mathbf{x}$. In the new coordinate system, the equation of σ simplifies to $\mathbf{y}^T \Lambda \mathbf{y} = 0$, which is nothing more than $\mathbf{y}_{3*}^T \Lambda_{33} \mathbf{y}_{3*} = 0$. The equation of the dual is $\mathbf{y}_{3*}^T \Lambda_{33}^{-1} \mathbf{y}_{3*} = 0$, $y_4 = y_5 = \dots = y_{N+1} = 0$.

Let $\mathbf{x} \in \mathcal{D}(\sigma)$ be a point in the dual (curve), written in the original coordinate system. Then $\mathbf{z} = M_{\mathcal{H}} \mathbf{x}$ is a point in $\bar{\sigma}$, the $\|\cdot\|$ -coords representation of σ . That is, $\mathbf{z} = M_{\mathcal{H}} U \mathbf{y}$, since U is unitary (actually orthogonal). Since all but the first three coordinates of \mathbf{y} vanish, we can just focus on $\mathbf{z} = M_{\mathcal{H}} U_{*3} \mathbf{y}_{3*}$. Substituting $\mathbf{y}_{3*} = (M_{\mathcal{H}} U_{*3})^{-1} \mathbf{z}$ into the equation of the dual, we obtain the equation of the $\|\cdot\|$ -coords representation of σ in homogeneous coordinates:

$$\bar{\sigma} : \mathbf{z}^T (M_{\mathcal{H}} U_{*3})^{-T} \Lambda_{33}^{-1} (M_{\mathcal{H}} U_{*3})^{-1} \mathbf{z} = 0.$$

11.3.5 A Numerical Example: An Elliptical Cone in 5-Space

It will be very helpful to have a matrix calculator such as octave or rlab at hand for verifying the computations in this section. Consider the hypersurface $\sigma : \mathbf{x}^T Q \mathbf{x} = 0$ in 5-space, where

$$Q = \begin{pmatrix} 32 & 0 & -16 & -80 & -64 & 8 \\ 0 & -1 & -1 & 2 & 1 & -2 \\ -16 & -1 & -17 & 114 & 81 & 6 \\ -80 & 2 & 114 & -20 & 14 & -52 \\ -64 & 1 & 81 & 14 & 31 & -38 \\ 8 & -2 & 6 & -52 & -38 & -8 \end{pmatrix}.$$

First let us verify that the rank of Q is 3 (using the calculator). In the notation of the duality section, here we have $N = 5$, $e = 0$, $f = 3$, $c = 1$.

Now diagonalize Q as $U \Lambda U^T$, where

$$U = \begin{pmatrix} 0.479 & -0.282 & 0.094 & 0.136 & -0.039 & 0.813 \\ -0.007 & 0.009 & -0.399 & 0.205 & 0.891 & 0.063 \\ -0.475 & 0.589 & -0.292 & -0.176 & -0.138 & 0.541 \\ -0.501 & -0.629 & 0.098 & -0.541 & 0.159 & 0.163 \\ -0.496 & -0.322 & -0.099 & 0.766 & -0.225 & 0.053 \\ 0.219 & -0.272 & -0.853 & -0.172 & -0.330 & -0.112 \end{pmatrix}$$

and

$$\Lambda = \begin{pmatrix} 201.3 & 0 & 0 & 0 & 0 & 0 \\ 0 & -178.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -6.241 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

In octave, calling the `eig()` function to compute the eigenvectors and eigenvalues of Q gives the desired result, except that the order has to be rearranged so that the nonvanishing eigenvalues appear at the upper left corner.

For the $\bar{\sigma}_{12345}$ curve, the transformation matrix for \mathcal{I} is

$$M_{\mathcal{I}} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 & 3 & 4 \\ -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

To find $\bar{\sigma} : [z, x, y]R[z, x, y]^T = 0$, the equation of the $\|\cdot$ -coords representation of σ , we compute $\bar{Q} = (M_{\mathcal{I}}U_{*3})^{-T}\Lambda_{33}^{-1}(M_{\mathcal{I}}U_{*3})^{-1}$ and find that

$$\bar{Q} = \begin{pmatrix} -10.573 & 3.662 & 11.776 \\ 3.662 & -1.268 & -4.076 \\ 11.776 & -4.076 & -13.112 \end{pmatrix}.$$

We need to find a simple parametric equation for this curve in order to draw its picture. To this end, we try to write \bar{Q} as

$$\begin{pmatrix} 1 & \mathbf{c}^T \\ 0 & V \end{pmatrix} \begin{pmatrix} f & 0 \\ 0 & L \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \mathbf{c} & V^T \end{pmatrix},$$

where V is a 2×2 rotation matrix and \mathbf{c} is the translation vector. This can be done by first diagonalizing the lower-right 2×2 block of \bar{Q} :

$$\begin{aligned} \bar{Q} &= VLV^T \\ &= \begin{pmatrix} 0.297 & -0.955 \\ 0.955 & 0.297 \end{pmatrix} \begin{pmatrix} -14.379 & 0 \\ 0 & -0.00133 \end{pmatrix} \begin{pmatrix} 0.297 & 0.955 \\ -0.955 & 0.297 \end{pmatrix}. \end{aligned}$$

Then $\mathbf{c} = L^{-1}V^T(3.662, 11.776)^T = (-0.858, 0.58)^T$. Finally, $f = -10.573 - \mathbf{c}^T L \mathbf{c} = 0.00426$. The matrix $\begin{pmatrix} 1 & 0 \\ \mathbf{c} & V^T \end{pmatrix}$ gives the necessary rotation and translation for simplifying the equation of $\bar{\sigma}$ in inhomogeneous equations. After a little further computation, we obtain the parametric equation as

$$\mathbf{x}(t) = \begin{pmatrix} 0.297 & -0.955 \\ 0.955 & 0.297 \end{pmatrix} \begin{pmatrix} 0.017 \cos(t) \\ 1.789 \sin(t) \end{pmatrix} + \begin{pmatrix} 0.809 \\ 0.647 \end{pmatrix}.$$

The same technique can be used to find a rigid motion (rotation + translation) that centers σ at the origin of the new coordinate system and places the axes of σ aligned with the new axes. Specifically, by decomposing $M_{\mathcal{H}}$ as $A^T B A$, where

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.340 & -0.010 & -0.663 & 0.616 & 0.294 & 0.307 \\ 0.137 & -0.403 & -0.347 & 0.092 & -0.119 & -0.833 \\ -1.250 & -0.831 & 0.054 & -0.281 & 0.375 & 0.295 \\ 3.438 & 0.384 & -0.281 & -0.511 & 0.681 & -0.222 \\ 0.577 & -0.010 & -0.599 & -0.521 & -0.543 & 0.274 \end{pmatrix}$$

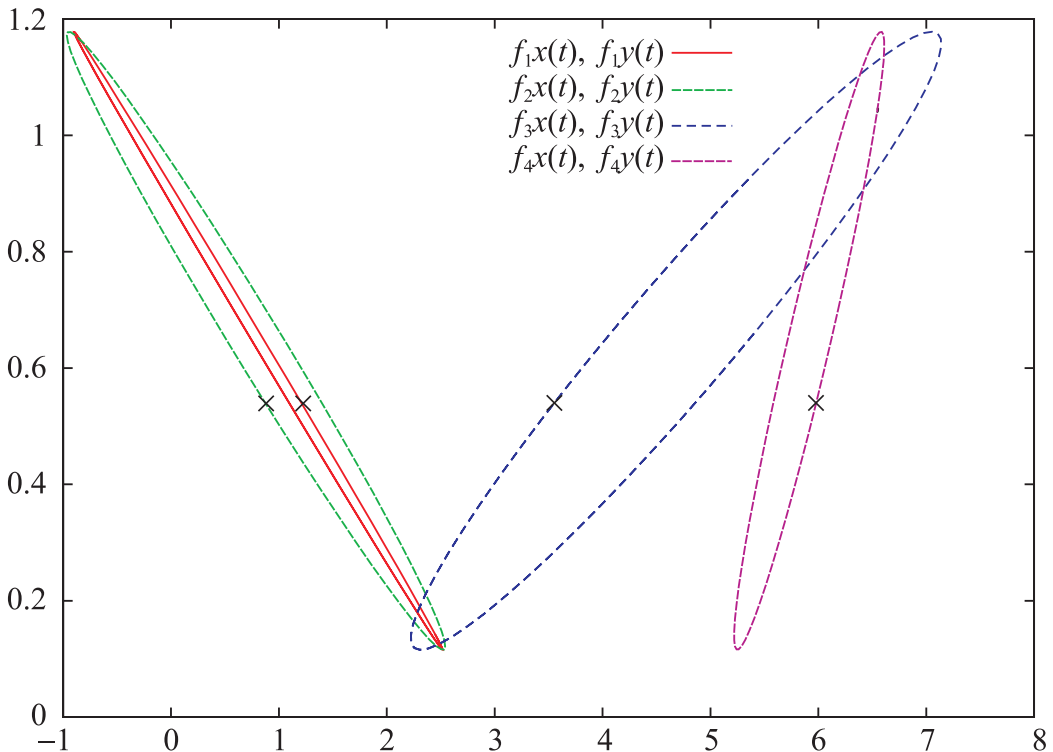


Figure 11.13. The $\|\text{-coord}$ representations of a cone in 5-space.

The indicated four points on the curves, with the same y coordinate, represent a plane tangent to the cone.

and

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -161.8 & 0 & 0 & 0 & 0 \\ 0 & 0 & -6.16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 153 \end{pmatrix},$$

we find that σ is an elliptical cone of the form $161.8x_1^2 + 6.16x_2^2 = 153x_5^2$. Figure 11.13 shows all four of the indexed curves representing σ , along with a set of indexed points representing $\pi_0 : x_1 + 4x_2 + 4x_3 - 4x_4 - 2.3258x_6 + 2 = 0$, a randomly chosen tangent plane of σ .

11.4 Network Visualization and Analysis with Parallel Coordinates

Yaron Singer

School of Computer Science,

Tel Aviv University, Israel

yaronsinger@yahoo.com

and

Ohad Greenshpan

IBM Research Center

Haifa, Israel

ohadg@il.ibm.com

11.4.1 Introduction

Large networks impose a challenge in many areas of research, and visualization plays a major role in analyzing topological traits and large amounts of (associated) information. Methods such as SeeNet [1], Genius [2], SpectralNet [3] and approaches suggested by Bolser et al. [4] and McGrath et al. [5] all display a variety of network visualization techniques in different research areas, dealing with complexity and data of large-scale networks. Although the intersection between the disciplines of network topological analysis and multivariate visualization is not immediate, with the wide spectrum of applications that we have thus far witnessed [6], [7], [8], it is clear that topological network analysis could benefit from multivariate visualization in general, and $\|\cdot\|$ -coords in particular. In this chapter

we introduce the Node Extraction Visualization (NEVIS) method, which enables network topological visualization using \parallel -coords.

Since networks are subject to changes that affect their function and performance, understanding the effects due to changing conditions is vital to their analysis. Here, we use changing conditions of node extractions to obtain topological insight about the network. Though the likelihood of such an event varies from one network to another, in instances where network functionality is determined largely by path distances and connectivity of nodes, node extraction is a crucial event. NEVIS transforms a given network into an equivalent representation in \parallel -coords. Under this transformation, networks under changing conditions are produced whose functionality is visualized. Since NEVIS transforms nodes into multidimensional points, we will see how it can be utilized to study key traits such as node stability, backup, inter-relationship between nodes, and to identify topological patterns in the network and other properties of interest in network research. We will also see that NEVIS's implementation is simple and modular.

11.4.2 The NEVIS Transformation

We begin by introducing and discussing *extracted networks* and the χ function, which are the basic elements of NEVIS. We will then see how to use these elements in the construction of the NEVIS algorithm.

Extracted Networks

Informally, dynamic networks are networks that are subject to topological change. We focus on a unique subset of dynamic networks, called *extracted networks*, which are defined as follows.

For the network modeled by a directed graph $S = (V, E)$ with nodes $V = \{v_1, v_2, \dots, v_n\}$ and edges $E = \{(v_i, v_j) \in V \times V | v_i \text{ is connected to } v_j\}$, let the i th *extracted network*, denoted by S_i , be the original network S without the node v_i and without any edges connected to it. That is,

$$S_i := (V \setminus \{v_i\}, E \setminus \{(v_j, v_k) \in V \times V | j = i \vee k = i\}). \quad (11.3)$$

For example, in Fig. 11.14, the network S and its third extracted network, S_3 , are shown. Transformation of a network with n nodes as above requires producing a set of extracted networks $\{S_1, S_2, \dots, S_n\}$. This set produces an n -dimensional space as explained in detail below.

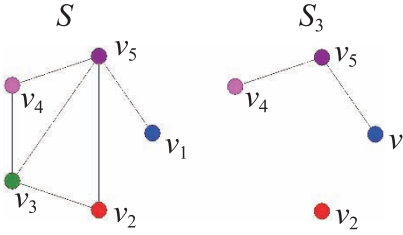


Figure 11.14. Network S and the third extracted network, S_3 .

Function χ : Measuring Node Centrality

Topological properties are frequently used to measure a network's functionality, which in turn depends on shortest paths and connectivity. To apply an appropriate measure that enables classification of nodes by these criteria, the χ function below is used to determine a node's topological centrality. For a network $S = (V, E)$, with V and E as above, and some node $u \notin V$, $\chi : V \cup \{u\} \rightarrow \mathbb{R}$ is defined by

$$\chi(v_i) = \begin{cases} \sum_{j=1, j \neq i}^n \frac{1}{d(v_i, v_j)}, & v_i \in V, \\ -\infty, & v_i \notin V, \end{cases}$$

with the following conventions:

1. $d(v_i, v_i) = 0 \forall i \in \{1, 2, \dots, n\}$,
2. $d(v_i, v_j) = \infty \iff v_j$ cannot be reached from v_i ,
3. if $d(v_i, v_j) = \infty$ then $\frac{1}{d(v_i, v_j)} = 0$.

Clearly, $0 \leq \chi(v_i) \leq n - 1 \forall v_i \in V$.

To further clarify, for the network shown in Figure 11.15, the χ function values for each node are given in Table 11.2.

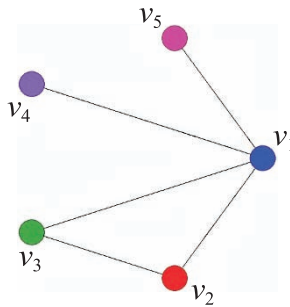


Figure 11.15. Network S whose χ function values are given in Table 11.2.

Table 11.2.

v_i	$d(v_i, v_1)$	$d(v_i, v_2)$	$d(v_i, v_3)$	$d(v_i, v_4)$	$d(v_i, v_5)$	χ value
v_1	0	1	1	1	1	$\sum_{j=1, j \neq 1}^n \frac{1}{d(v_1, v_j)} = 4$
v_2	1	0	1	2	2	$\sum_{j=1, j \neq 2}^n \frac{1}{d(v_2, v_j)} = 3$
v_3	1	1	0	2	2	$\sum_{j=1, j \neq 3}^n \frac{1}{d(v_3, v_j)} = 3$
v_4	1	2	2	0	2	$\sum_{j=1, j \neq 4}^n \frac{1}{d(v_4, v_j)} = 2.5$
v_5	1	2	2	2	0	$\sum_{j=1, j \neq 5}^n \frac{1}{d(v_5, v_j)} = 2.5$

Why χ ? For a node $v_i \in V$, note that $\chi(v_i)$ depends on the number of nodes connected to v_i and their distances from it; χ monotonically increases with respect to both centrality and connectivity of the node. Thus, in relation to other nodes in the network, for a node v_i that is connected to a large number of nodes at short distances, $\chi(v_i)$ is high; symmetrically, for a node v_j that is connected to a small number of nodes at large distances, $\chi(v_j)$ is low. These properties make the χ function a favorable candidate for measuring a node's centrality in the network.

The NEVIS Algorithm

Having introduced its elements, we now present the NEVIS transformation and the algorithm used to implement it.

For a network S as described above, we produce n extracted networks S_1, S_2, \dots, S_n . By extracting the node v_i from S , we produce the i th extracted network S_i , which consists of $n - 1$ nodes. For each extracted network, using the breadth-first search (BFS) algorithm for graphs to obtain edge distances between nodes, we calculate the χ function values of all n nodes (by definition, $\chi(v_i) = 0$ in $S_i \forall i \in \{1, \dots, n\}$). This process produces a matrix of order $n \times n$:

CreateMatrix

```

for all  $v_i \in V$  do
   $S_i \leftarrow \text{ExtractNode}(v_i, S)$ 
  for all  $v_j \in V \setminus \{v_i\}$  do
     $\text{BFS}(S_i, v_j)$ 
     $\chi(v_j) \leftarrow 0$ 
    for all  $v_k$  reachable from  $v_j$  do
       $\chi(v_j) \leftarrow \chi(v_j) + \frac{1}{d(v_j, v_k)}$ 

```

```

    end for
     $M_{ij} \leftarrow \chi(v_j)$ 
  end for
end for
return  $M$ 

```

The matrix is then transformed to its equivalent multidimensional representation by considering each row as an n -dimensional point. With parallel coordinates, each point is represented as a polygonal line. We obtain visualization with the networks S_1, S_2, \dots, S_n represented by the multidimensional coordinates system with first, second, \dots , n th axes respectively. In turn, the nodes v_1, v_2, \dots, v_n are represented by n polygonal lines. Figure 11.16 describes a flow of this implementation.

NEVIS transforms a network with n nodes into n points in an n -dimensional space. The n extracted networks serve as the components of the n th dimensional space, and the nodes' χ values are transformed into points in that space. The following section discusses the properties of this transformation and their use in network analysis.

General Network Transformations

Note that we can generalize the idea of network transformation beyond NEVIS. We can formally categorize network transformations into parallel coordinates with the following functions:

1. γ function, which is defined on the set of all possible directed graphs and returns a set of networks.
2. ν function, which is defined on the set of all possible directed graphs and returns a real value.

In NEVIS we use the γ function, which produces the set of all possible extracted networks. That is, $\gamma(S) = \{S_1, S_2, \dots, S_n\}$ for a network with n nodes. The ν function used can be seen as a combination of ψ , which is defined by $\psi(S) = V$, and χ , as shown above. Thus selection of γ, ν, ψ, χ can yield different network transformations as desired.

11.4.3 Properties of NEVIS

We can now begin exploring some of the properties of NEVIS. The principal properties discussed here facilitate the analysis of large-scale networks, as presented later.

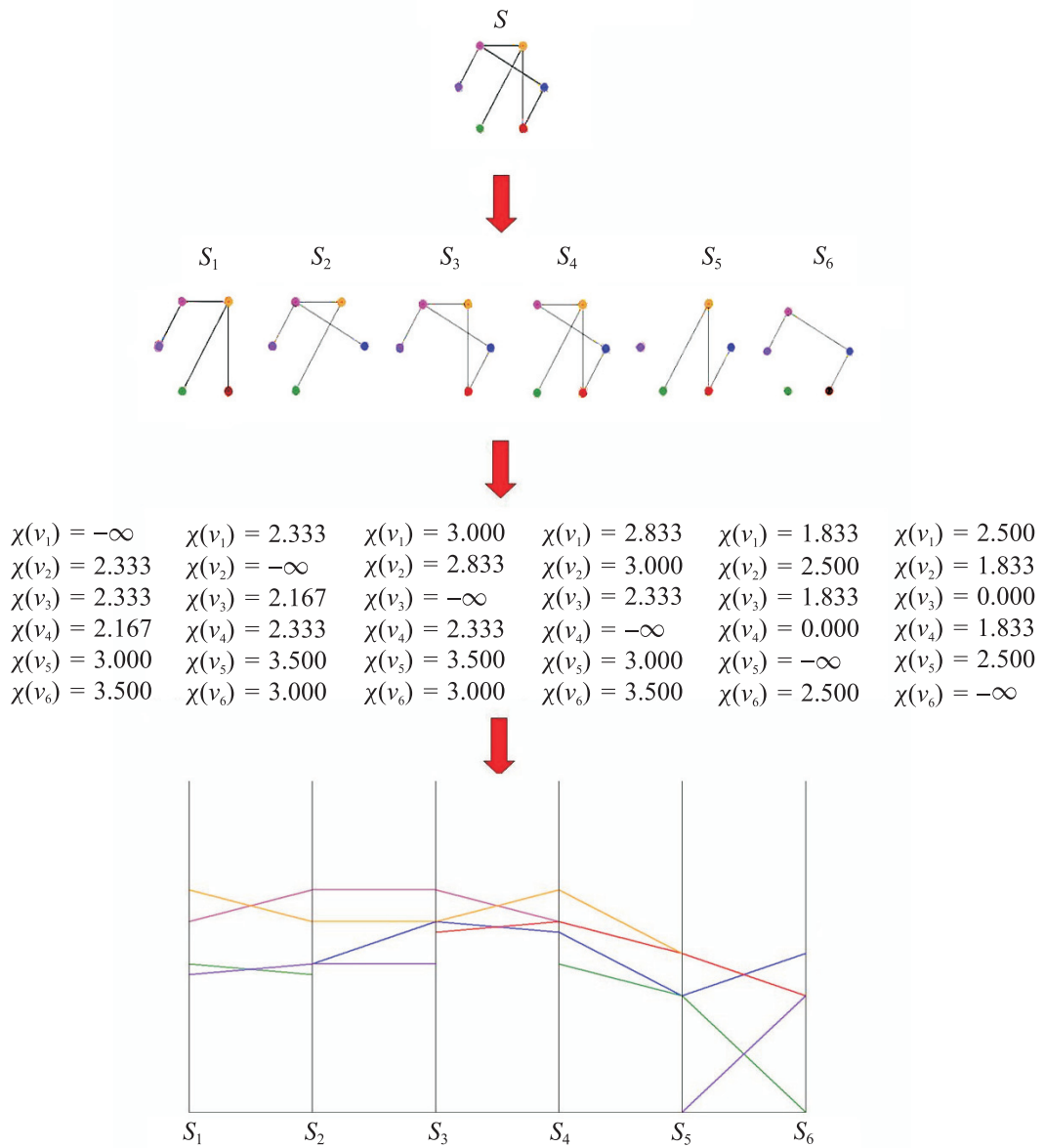


Figure 11.16. The NEVIS transformation network $\rightarrow n$ points in n -dimensional space.

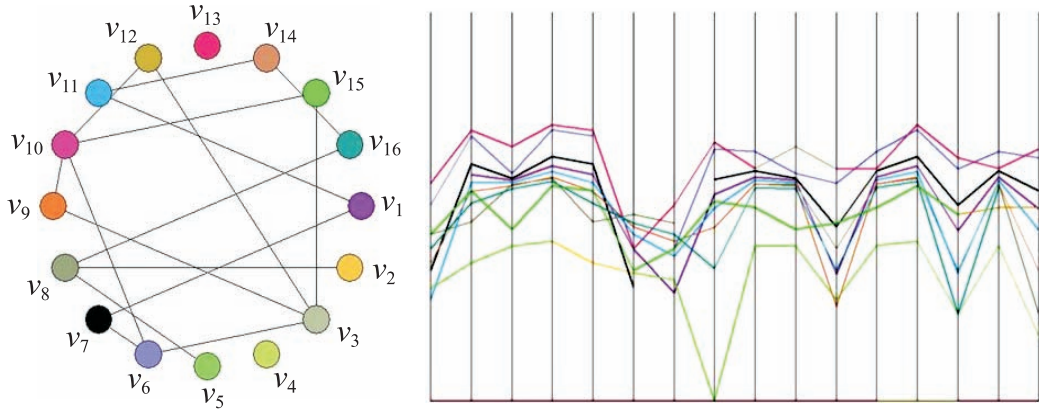


Figure 11.17. Visualizing topological significance.

The drops in the sixth axis as opposed to the consistency in the third axis suggest that v_3 is backed up in the network.

Topological Significance

After creating the n extracted networks S_1, S_2, \dots, S_n , the effects of each extraction are independently measured by $\chi(v_j) \forall v_j \in V \setminus \{v_i\}$. By studying the i th component of the n -tuple produced by NEVIS, one can comprehend the effect of extracting v_i from the network. If v_i has a significant effect on the network's centrality and connectivity, this property is manifested in a drop of polygonal lines representing nodes depending on v_i in the i th axis in NEVIS. It thus follows that the topological significance of v_i correlates with the number of polygonal lines dropping in NEVIS, and the sharpness of their drop.

Contrary to our intuition, a high χ does not necessarily guarantee topological significance. Figure 11.17 displays an example in which v_3 has a high χ value in the network (this can be seen in NEVIS by the fact that its corresponding polygonal line is the top outlier), though observing the third axis in NEVIS, it is clear that v_3 's extraction is not significant to the network. On the other hand, there is a dramatic drop in the sixth component, despite the fact that $\chi(v_6)$ is not among the highest in the network. This seemingly counterintuitive example reveals one of the strong properties in NEVIS. A closer study of the above example shows that the extraction of v_6 is more significant to the network than that of v_3 's despite $\chi(v_6) < \chi(v_3)$, since each node connected to v_3 is also connected to v_{10} .

Intuitively, the idea of a *backup* of a node is the existence of alternative paths connecting a pair of its neighbors without passing through the node. In Figure 11.17, v_3 's extraction is not significant to the network's stability, since v_{10} serves as a full

backup. We can think of v_3 as having a backup of order 4 (v_3 's degree) in the sense that there is another node (v_{10}) connected to all of v_3 's four neighbors ($\{v_6, v_7, v_9, v_{10}\}$). Formally, suppose that v_i is connected to $\{v_1, \dots, v_r\}$. We say that v_i has a *backup of order k* , $k \leq r$, if there is at least one node in the network connected to k of v_i 's neighbors. Clearly a node's significance in the network highly depends on its backup order. Study of the drops in the i th axis in NEVIS provides a measure of v_i 's global significance which considers its backup order.

The ability to measure node significance considering both centrality and its backup order enables evaluating a node's impact on the network's topology in the network. Since a node can be partially backed up by a set of nodes that are reached at different distances, the ability to consider a complicated factor such as a node's backup makes NEVIS a useful tool.

Relationship between Nodes

An immediate conclusion from the above is that relationship between nodes v_j and v_k can be studied by examining the drop of v_k in the j th axis, and symmetrically examining the drop of v_j in the k th axis. Although the latter is one of its fundamental properties, NEVIS's representation of nodes as polygonal lines enables greater study of the relationship between nodes.

Since each node's χ value throughout n extractions is represented as a polygonal line, studying the relationship between the polygonal lines representing v_j and v_k provides insight into the relationship between these nodes. Consider the network portrayed in Figure 11.18, composed of three separate connected components. To achieve clarity, nodes in the same connected component are given an identical color. Assume that the axes are ordered in the manner displayed (we refer to this

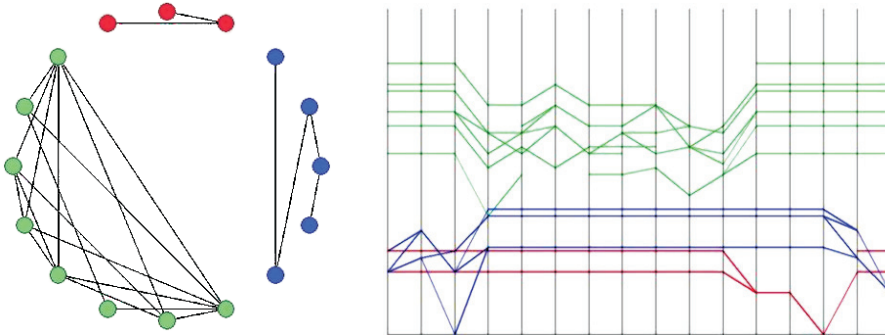


Figure 11.18. Network composed of strongly connected components. The figure displays the interrelationship between nodes.

point below). In this example it is evident that polygonal lines of the same color have a specified range in which they are subject to change and a complementary range where they run horizontally. It is easy to see that this property directly derives from the fact that nodes in the same connected component react similarly to extractions in the network, influenced only by extractions in the connected component to which they belong, and insensitive to extractions from other connected components. In this example we have chosen a specific axis ordering obtained by sorting according to χ value.

Stability of Nodes

Since the nodes' χ values are represented by polygonal lines in the NEVIS representation, peaks of polygonal lines disclose the corresponding nodes' centrality values in instances in which the node extracted has minimal effect. Note that for a network $S = (V, E)$ as above, it is possible to implement NEVIS as n points in an $(n + 1)$ -dimensional space by adding the 0th extracted network to S and thus obtaining visualization of the χ value in S itself. In Figure 11.19, the top outlier is shown by a red polygonal line, which represents v_{15} in the network, with noticeably higher values than the rest of the nodes in the network. This can be easily verified by observing the node-link diagram. NEVIS orders the polygonal lines, starting from the top, in order of decreasing centrality throughout all extractions.

Since the axes represent the extracted networks, it follows that polygonal lines free of dramatic inclines and declines represent nodes that are relatively stable

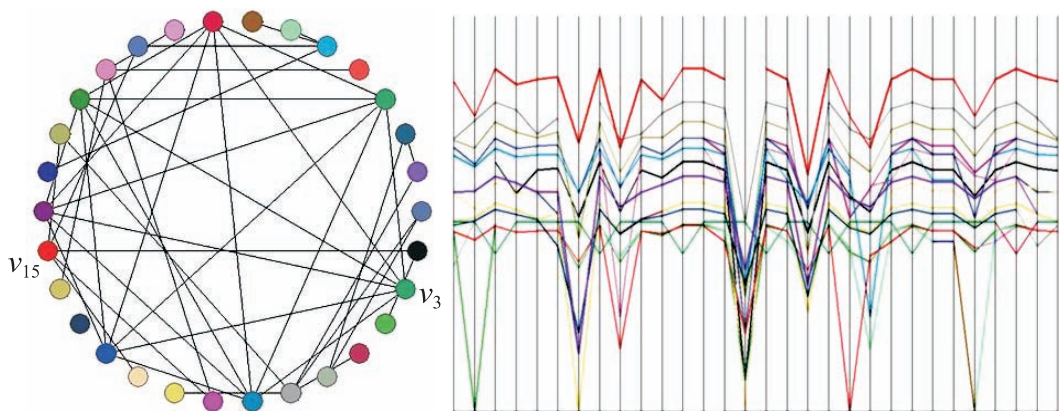


Figure 11.19. Analyzing stability from the behavior of polygonal lines. The green polygonal line remains relatively steady, whereas the red polygonal line has sharp drops.

and immune to node extractions in the network. Note that node stability does not necessarily correlate with high centrality value. In Figure 11.19, the sharp inclines and declines of the red polygonal line testify that its corresponding node v_{15} is unstable in the sense that its centrality is highly affected by extractions. In contrast, it can be inferred that v_3 is relatively stable by inspecting the yellow polygonal line by which it is represented in NEVIS. In this case, it is not clear how this can be deduced by studying the node-link diagram alone. The ability to analyze node stability in the network can be widely used in network analysis, since it discloses weak links in the network.

Identifying Topological Patterns

Identifying unique topological patterns can be used in network analysis to deduce relevant information in a first encounter with a given network. To identify unique topological patterns, we use a slight variation of NEVIS by replacing the χ function with χ_0 defined by

$$\chi_0(v_i) = \begin{cases} \sum_{j=1, j \neq i}^n \frac{1}{d(v_i, v_j)}, & v_j \in V, \\ 0, & v_j \notin V. \end{cases}$$

NEVIS reveals the network's distinct topological properties by mapping them one-to-one into distinct patterns. Figure 11.20A displays a clique, a graph in which all the nodes are connected between one another, and Figure 11.20B shows a star, a set of independent nodes all connected to a single node.

Following a one-to-one transformation between topological properties and patterns in \parallel -coords, networks with topological patterns that are *almost* symmetric are represented as their *almost* symmetrical equivalents in NEVIS. Figures 11.21A and 11.21B display networks that are *almost* a clique and *almost* a star, respectively. Comparing the NEVIS representation of the *perfect* graphs with the *almost* perfect ones illustrates this point. This property is very useful in the study of topological properties in networks. Perfect topological patterns are rare; it is their *almost* counterparts which are encountered in the real world. We now show how this property can be used in network study.

11.4.4 Large-Scale Networks

In this section we discuss applications of NEVIS on large-scale networks. Since it synthesizes network analysis and multivariate data visualization, NEVIS benefits from both worlds; preprocessing algorithms that enable combinatoric reduction of the network can be applied as well as any methods that enable dealing with large

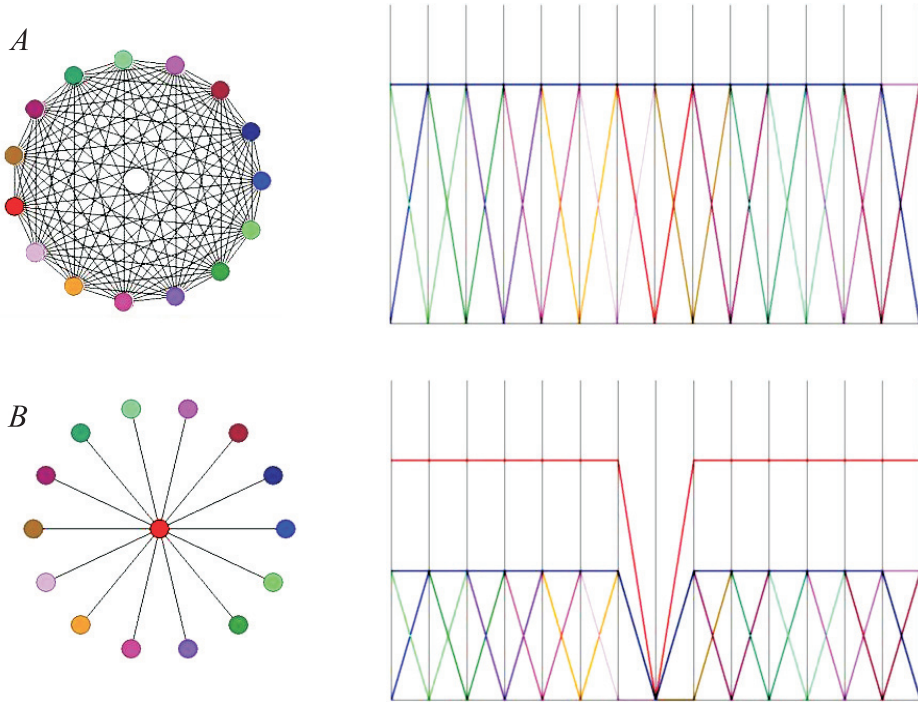


Figure 11.20. *Clique with 15 nodes is in A, and a star with 15 nodes is in B.*

data sets in $\|$ -coords. We display examples of both in this section. Our interest here is not to offer novel methods for re-scaling networks or reducing large data sets for visualization, but rather to show how existing methods of these ideas can be applied in NEVIS.

Matrix Manipulations

We can now view possible simple solutions to deal with large-scale networks utilizing $\|$ -coords. The need to analyze large networks raises the issue of scalability, which can be accomplished with $\|$ -coords. Figure 11.22A displays a node-link diagram of a network that consists of 100 nodes, and its NEVIS presentation is represented in Figure 11.22B. Due to NEVIS's simple construction, selecting a subset of nodes to be visualized is easy. Recall from Section 11.4.2 that implementing NEVIS requires producing a matrix that represents the nodes' χ values throughout extractions. To ignore polygonal lines, we simply replace the values of their corresponding rows in the produced matrix with $-\infty$. That is, in order to ignore

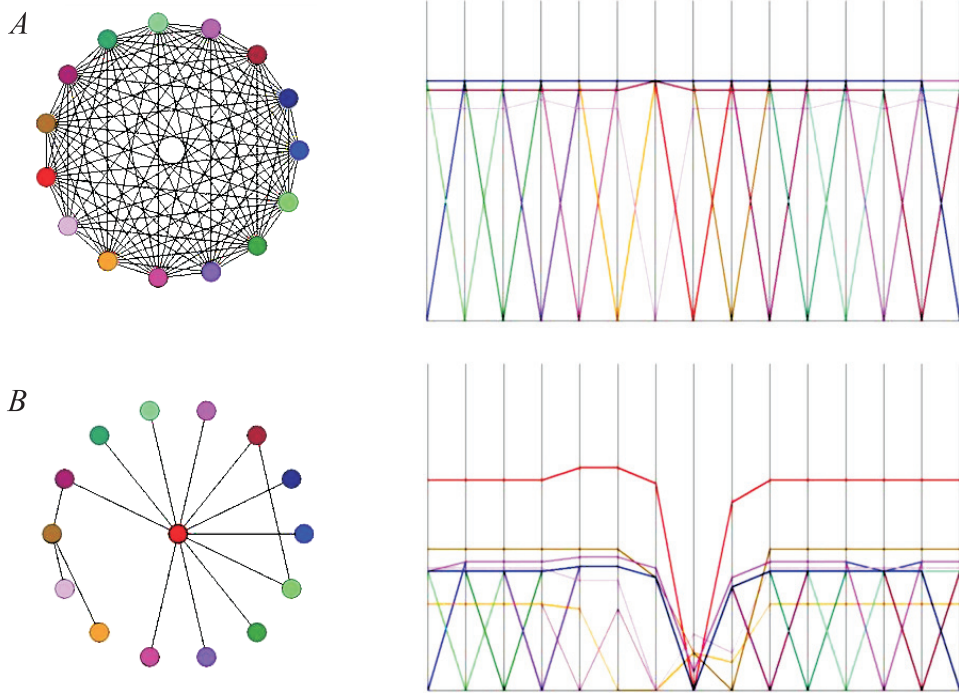


Figure 11.21. In A, a network that is *almost* a clique, and in B an *almost* star topology.

the v_j in the NEVIS visualization, replace the j th row in the matrix returned by `CreateMatrix`:

$$\begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \vdots & \vdots \\ x_{j1} & \dots & x_{jn} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nn} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ -\infty & \dots & -\infty \\ \vdots & \vdots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \vdots & \vdots \\ -\infty & \dots & -\infty \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nn} \end{pmatrix}.$$

The results of this procedure are shown in Figure 11.22C. Similarly, it is easy to reduce a subset of axes from NEVIS by reducing the matrix to represent the desired axes only. An example of such reduction is shown below. The result of reducing both polygonal lines and axes is displayed in Figure 11.22D and corresponds to the matrix

$$\begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1n} \\ x_{21} & \dots & x_{2j} & \dots & x_{2n} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{n1} & \dots & x_{nj} & \dots & x_{nn} \end{pmatrix} + \begin{pmatrix} 0 & \dots & -\infty & \dots & 0 \\ 0 & \dots & -\infty & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & -\infty & \dots & 0 \end{pmatrix} \\
 = \begin{pmatrix} x_{11} & \dots & -\infty & \dots & x_{1n} \\ x_{21} & \dots & -\infty & \dots & x_{2n} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{n1} & \dots & -\infty & \dots & x_{nn} \end{pmatrix}.$$

The running time required to implement NEVIS depends on the CreateMatrix procedure, and visualization is thereafter achieved in linear time. This implementation allows performance of such manipulations efficiently. Moreover,

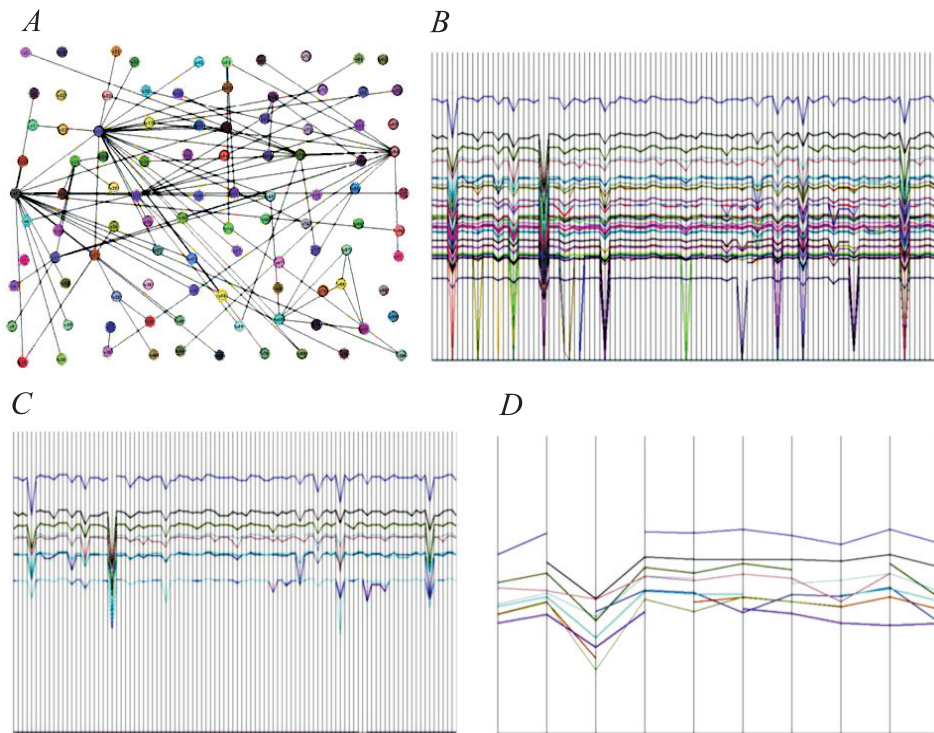


Figure 11.22. Examples of possible solutions for large networks. Shown are polygonal line reductions that represent 10 nodes in a 100-dimensional space in C, and D shows the 10 nodes in their corresponding 10-dimensional space.

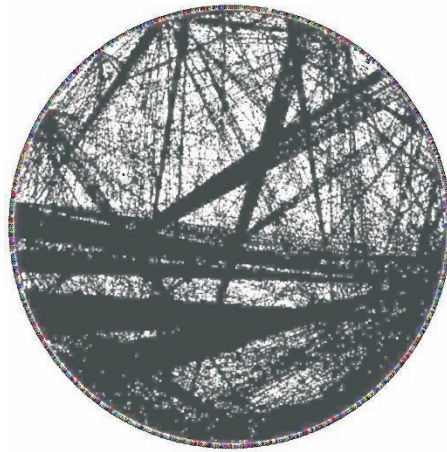


Figure 11.23. NEVIS applied to a genetic regulatory network with over 3000 nodes. Only one-fifth of the nodes are shown.

due to this implementation, a wide range of solutions to dealing with large data sets in \parallel -coords can be applied in a simple manner.

Combinatoric Manipulations

Figure 11.23 displays a node-link diagram of a genetic regulatory network of over 3000 nodes. The network shown is a protein regulatory network, modeled by a directed graph, where proteins are represented by nodes and regulatory interactions are represented by edges. Clearly, the size of the network raises serious obstacles in attempts for visual analysis. To obtain visualization of the above network, we apply NEVIS to the network and show nodes with out-degree greater than zero. The reasoning here is that nodes with zero out-degree have little effect on χ values (why?), and visualization in NEVIS does not provide any relevant information. We thus obtain a visualization of 142 nodes of the network in NEVIS, as shown in Figure 11.24.

We can now use the properties of NEVIS from the section above to analyze the network. In this instance a few immediate conclusions can be drawn about the network. Generally, apart from a few specific instances, the polygonal lines run relatively horizontally, which testifies that the network is relatively stable. Note also that there is an area that is clear of polygonal lines. We conclude that no node receives χ values in that area, and that it may testify to segmentation of two separate connected components in the network. The horizontal polygonal lines in the lower values support this hypothesis, similar to the connected-components

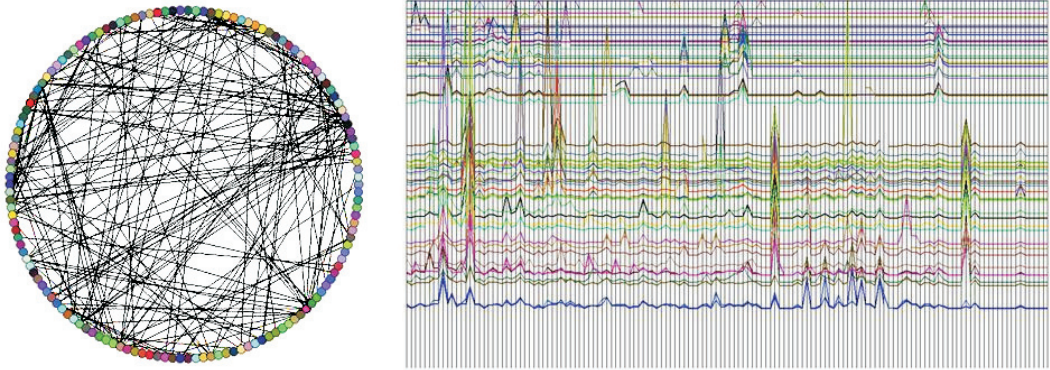


Figure 11.24. Visualization of the 142 nodes in the regulatory network of Figure 11.23 with degree greater than zero.

example presented in the previous section. Examining the top blue polygonal line, it is evident that despite its high centrality value, its extraction does not affect the network's functionality. Using an interactive tool that enables pointing on the polygonal line and obtaining its number, we can inspect its extraction effect on other nodes. Examining the 127th axis, we learn that its extraction does not affect functionality in the network. Closer study reveals that there are, in fact, two polygonal lines of maximal χ values that race between one another throughout all extraction. We assume that this node serves as its backup.

There are four nodes whose extraction has significant effect on the network. This is manifested in the drops of polygonal lines in the area of the 9th, 15th, 34th, 82nd, 124th axes. This signifies that these nodes have relatively poor backup in the network, and their extraction affects the network's functionality.

This example is applicable to biological networks, the Internet and elsewhere. Other works on combinatoric reduction of the problem can also be applied as preprocessing algorithms in a similar manner to reduce the combinatoric problem.

Exercises

1. What is the running time of the CreateMatrix algorithm?
2. Give an example of a *connected* graph that has horizontal portions in NEVIS.
3. Suppose that for $S = (V, E)$ and some node $u \notin V$ we implement NEVIS, replacing the χ function with the function $\deg : V \cup \{u\} \rightarrow \mathbb{N}$ defined by

$$\deg(v) = \begin{cases} |\{w : w \text{ is connected to } v\}|, & v \in V, \\ -\infty, & v \notin V. \end{cases}$$

- (a) Apply NEVIS with the \deg and χ (not χ_0) functions on a five-node star and five-node clique. Any difference? Why?
 - (b) What would be the motivation to replace \deg with χ for topological pattern identification?
 - (c) Think of other topological functions under which a star would look the same as in NEVIS.
4. This question discusses backup order, which has been defined above.
- (a) Prove that the relation $R_k = \{\langle u, v \rangle \mid u \text{ is a backup of order } k \text{ of } v\}$ is symmetric. That is, prove that $u \text{ is a backup of order } k \text{ of } v \iff v \text{ is a backup of order } k \text{ of } u$.
 - (b) Is it possible that u has a backup of order k , where $k = \deg(u)$, and there will still be a significant drop in its axis?

References

- [1] Becker, R.A., Eick, S.G., Wilks, A.R., Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics* (1995), Vol. 1–1; 16–21.
- [2] Ciccicarese, P., Mazzocchi, S., Ferrazzi, F., Sacchi, L., Genius: A new tool for gene networks visualization. *Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP) Proceedings* (2004).
- [3] Forman, J.J., Clemons, P.A., Schreiber, S.L., Haggarty, S.J., SpectralNET—an application for spectral graph analysis and visualization. *BMC Bioinformatics*. 2005 October 19; 6:260.
- [4] Bolser, D., Dafas, P., Harrington, R., Park, J., Schroeder, M., Visualisation and graph-theoretic analysis of a large-scale protein structural interactome. *BMC Bioinformatics* 2003; 4:45.
- [5] McGrath, C., Hardt, D.C., Blythe, J., Visualizing Complexity in networks: Seeing both the forest and the trees. *Connections* 25(1):37–41 2002 INSNA.
- [6] Keim, D.A., Information Visualization and Visual Data Mining. *IEEE TVCG*, 7–1, 2002, pp. 100–107.
- [7] Shneiderman, B., The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings of the 1996 IEEE Symposium on Visual Languages*.
- [8] Wegenkittl, R., Loffelmann, H., Groller, E., Visualizing the Behaviour of Higher Dimensional dynamical systems. *Proceedings of IEEE VIS* 1997 pp. 119–126.

11.5 To See \mathbb{C}^2

Yoav Yaari
School of Mathematical Sciences
Tel Aviv University
Tel Aviv, Israel
 yoav.yaari@gmail.com

11.5.1 The Complex Field \mathbb{C}

The set of complex numbers $x + iy$, where x and y are real numbers, together with the binary operations $(+, \times)$, forms the field \mathbb{C} . Using $(1, i)$ as a basis, \mathbb{C} can be identified with the two-dimensional plane \mathbb{R}^2 , the horizontal and vertical axes being the real and imaginary axes, respectively. Addition corresponds to translation, while multiplication involves scaling (of the absolute values) and rotation (Fig. 11.25).

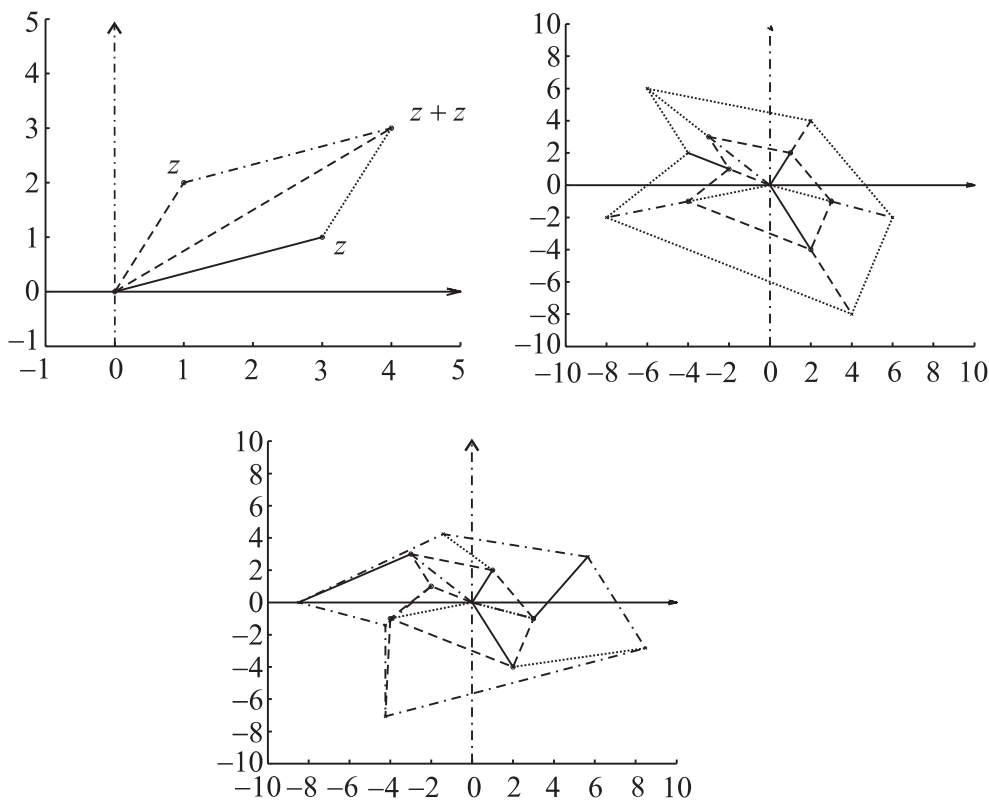


Figure 11.25. Translation, scaling, and rotation.

The goal here is the visualization of \mathbb{C}^2 , the space of complex-valued functions, which ostensibly is four-dimensional. Yet trying to represent \mathbb{C}^2 with four parallel axis does not seem possible even for a simple function like $f = az + b$. No axis assignment (i.e., choosing which axis should represent the real and imaginary parts) or subsequent axis ordering worked.

11.5.2 The Space \mathbb{C}^2

Though we are dealing with four-dimensional information, rather than four independent coordinates, what is involved are *pairs of pairs*. This suggests using two copies of \mathbb{C} , which we take as parallel planes with a horizontal one-dimensional axis w passing through their origins. Of course, this arrangement collapses to $\|$ -coords when the imaginary parts are zero.

The two copies of \mathbb{C} are denoted by $Z_i, i = 1, 2$, with X_i, Y_i as the real and imaginary axis, respectively. The origin of Z_1 is at $w = 0$ and the origin of Z_2 at $w = 1$; the distance between the parallel planes is one unit (Fig. 11.26). The

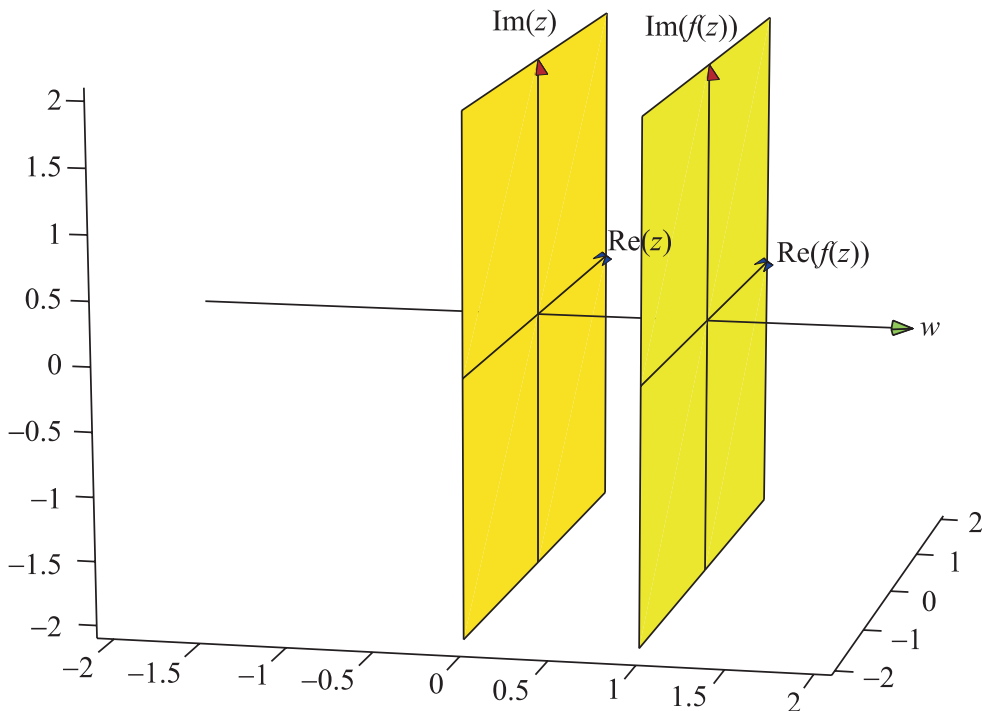


Figure 11.26. The complex parallel planes coordinate system.

coordinates (x, y, w) designate the position of a point in the three-dimensional space \mathbb{R}^3 . For example, $(2, 1, 0)$ stands for the point $2i + 1$ on Z_1 , and $(1, -2, 1)$ for the point $i - 2$ on Z_2 . To visualize (i.e., represent) the pair $\langle z_1, z_2 \rangle \in \mathbb{C}^2$ we propose joining the point $z_1 \in Z_1$ to $z_2 \in Z_2$ with a *straight line*. Taking our cue from the $\|\text{-coords}$ *point-line* duality, we seek a corresponding relation resulting from the visualization of \mathbb{C}^2 .

Let's see what the representation of a simple linear function, say $f(z) = 2z$, looks like. The collection of lines joining the points $z \in Z_1$ to $2z \in Z_2$ is given parametrically by $(\Re(z + tz), \Im(z + tz), t) t \in \mathbb{R}$. They intersect at $(0, 0, -1)$ (Fig. 11.27), which is an auspicious result. Recall that for $f(x) = ax + b$ $a, b \in \mathbb{R}$, the representing point in $\|\text{-coords}$ is $(\frac{1}{1-a}, \frac{b}{1-a})$. Proceeding in the same way for the function $f(z) = az + b$ with $a \in \mathbb{R}$ and $b \in \mathbb{C}$, the representing point obtained is $(\Re(\frac{b}{1-a}), \Im(\frac{b}{1-a}), \frac{1}{1-a})$. Just as for $\|\text{-coords}$, the representing point is undefined in \mathbb{R}^2 when $a = 1$, prompting the extension of the representation from \mathbb{R}^2 to \mathbb{P}^2 .

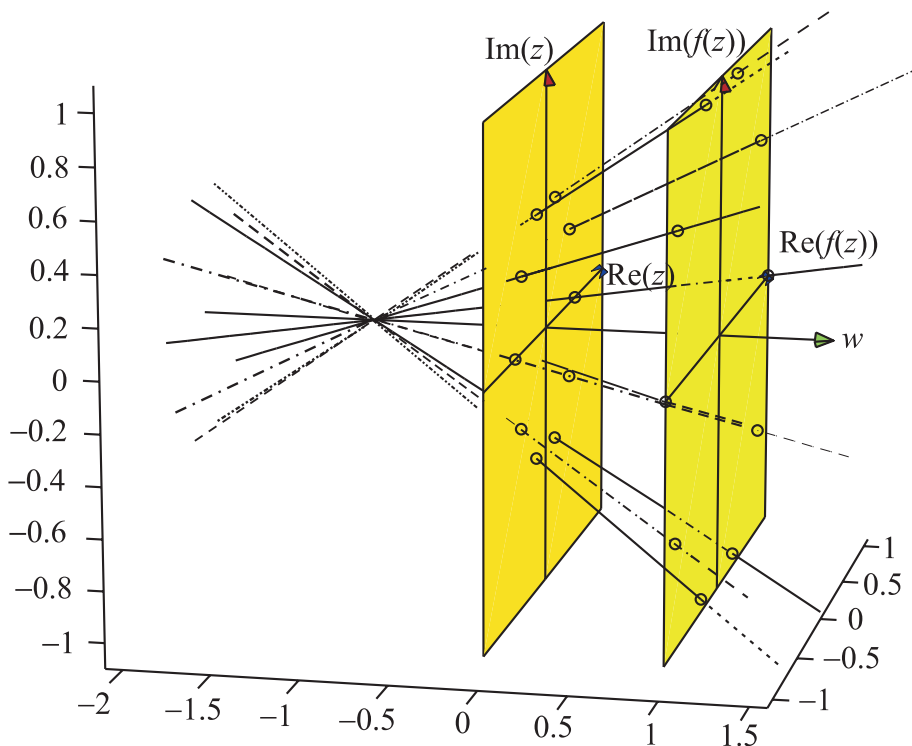


Figure 11.27. The lines joining the points $(z, 2z)$ intersect a point.

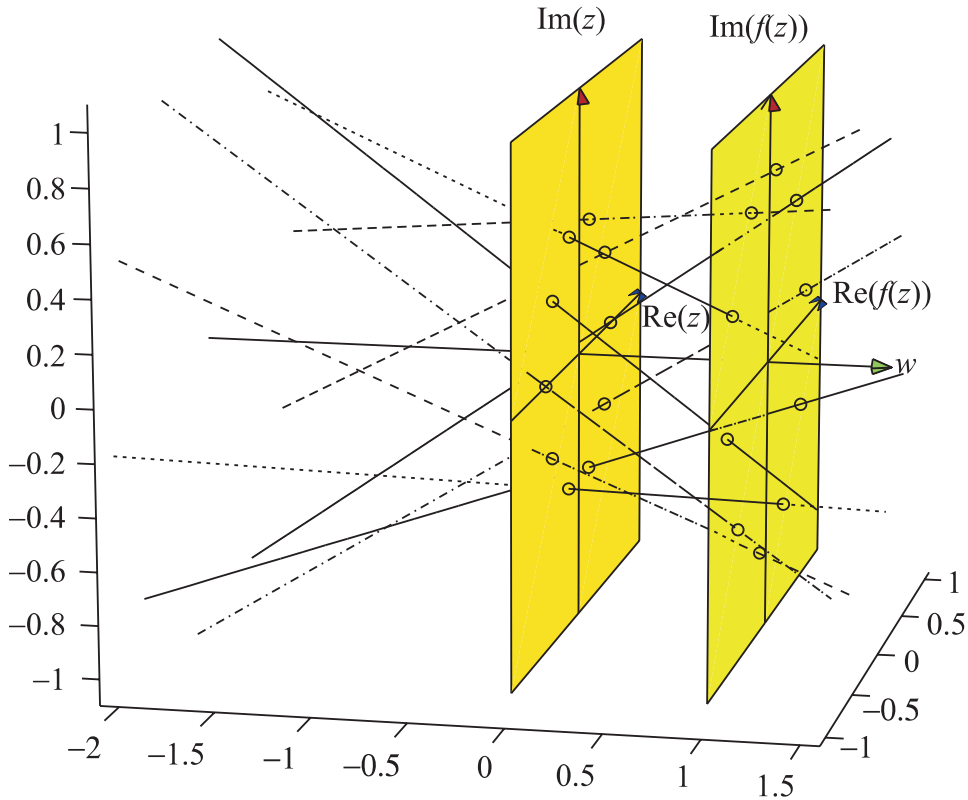


Figure 11.28. The lines joining the points $(z, (1+i)z)$ do not intersect.

For the representation of $f(z) = az$, where a is a complex number (but not real), the lines joining the pairs $\langle z, az \rangle$ are $(\Re(z + taz - tz), \Im(z + taz - tz), t)$. To find their intersection, we look for points $p, q \in Z_1$ ($p \neq q$) such that $p + tap - tp = q + taq - tq$, with t being their common w coordinate. That is, $1 + ta - t = 0$ or $t = \frac{1}{1-a}$, but t must be real; therefore, these lines *do not intersect* for complex a , as shown in Figure 11.28.

11.5.3 Representing Points on a Line in \mathbb{C}^2

Showing the association between z in Z_1 with $f(z)$ in Z_2 with straight lines is abandoned, since it does not yield a duality for a linear function with complex coefficients. Can a duality be obtained if z and $f(z)$ are joined by a *curve*? Let us return to the function $f(z) = az + b$, and observe that it consists of a scaling

by $|a|$, rotation of a radians, and translation by b . Rewriting the function in the form

$$f(z) = az + b = az + b \frac{1-a}{1-a} = a \left(z - \frac{b}{1-a} \right) + \frac{b}{1-a},$$

reveals that the point $(\Re(\frac{b}{1-a}), \Im(\frac{b}{1-a}))$, which we call the *pivot*, plays an important role. We seek curves that reduce to straight lines for $a \in \mathbb{R}$, to preserve the duality which holds in this case. Let us describe the curves by the function

$$h(z, w) = g(a, w)(z - p) + p, \quad p = \frac{b}{1-a},$$

which is essentially the same as $f(z) = az + b$. The coefficient a is replaced by $g(a, w)$ to distribute the rotation and scaling along the w axis. In time we will remove the dependence of h on z . The additional constraints are:

1. $g(a, 0) = 1$, so the curve contains $(\Re(z), \Im(z), 0)$, and
2. $g(a, 1) = a$, so the curve contains $(\Re(az + b), \Im(az + b), 1)$. Then the curve joins the points z in Z_1 to $az + b$ in Z_2 .

We would like the distance from the pivot to points on the curve given by h to vary linearly with w . By the way, this also allows us to stay “close” to $\|\cdot$ -coords. Incorporating the two conditions above we obtain:

$$|g(a, w)| = 1 + (|a| - 1)w. \quad (11.4)$$

In addition, the first two boundary values suggest an exponential function $g(a, w) = a^{L(a, w)}$, and computing $L(a, w)$ from (11.4) yields

$$L(a, w) = \frac{\log(1 + (|a| - 1)w)}{\log(|a|)}, \quad |a| \neq 1.$$

Although the function is undefined when $|a| = 1$, it is easy to see that it has a limit as $|a| \rightarrow 1$, and we can set $L(1, w) = w$. This function satisfies the criteria though it does not have a value for every w . As we know, the *log* function is defined only for positive arguments. We conclude that

$$|a| > 1 \Rightarrow L(a, w) \text{ is defined only for } w > \frac{1}{1 - |a|},$$

$$|a| < 1 \Rightarrow L(a, w) \text{ is defined only for } w < \frac{1}{1 - |a|},$$

$|a| = 1 \Rightarrow L(a, w)$ is defined for all w .

These conditions divide the w axis in three portions. As will be seen in the subsequent examples, the open interval $(0, 1)$ between the parallel planes will be excluded from the representation of \mathbb{C}^2 . Altogether, then, the curve we obtained is given parametrically by:

$$\left(\Re \left(a^{\frac{\log(1+(|a|-1)w)}{\log(|a|)}} \left(z - \frac{b}{1-a} \right) + \frac{b}{1-a} \right), \right. \\ \left. \Im \left(a^{\frac{\log(1+(|a|-1)w)}{\log(|a|)}} \left(z - \frac{b}{1-a} \right) + \frac{b}{1-a} \right), w \right). \quad (11.5)$$

A visual examination of the curve reveals that it is a spiral advancing over the w -axis exponentially. It can also be described as a logarithmic spiral advancing over the w axis linearly. It originates at a pivot point on the line $(\Re(\frac{b}{1-a}), \Im(\frac{b}{1-a}), w)$ and it is on the surface of a cone whose apex is at the same pivot. To obtain this result we need find a w such that

$$a^{\frac{\log(1+(|a|-1)w)}{\log(|a|)}} = 0.$$

This occurs only when the exponent is $-\infty$ for $|a| > 1$ or ∞ for $|a| < 1$. Both of these conditions are satisfied as $w \rightarrow \frac{1}{1-|a|}$. Hence, the cone's apex, which is also this limit point of the spiral, is

$$\left(\Re \left(\frac{b}{1-a} \right), \Im \left(\frac{b}{1-a} \right), \frac{1}{1-|a|} \right). \quad (11.6)$$

This point is independent of z , so it must be the intersection of the all the curves representing points on \mathbb{C}^2 for the function $f(z) = az + b$. To complete the argument we need to verify that it is the *only* point of intersection. Recall that the w variation is *only* along the w axis and so the value of w is the same for both r, q which are on Z_1 . The two curves, one joining the pair $\langle r, ar + b \rangle$ and the other joining $\langle q, aq + b \rangle$ for $r \neq q$, are subtracted to obtain

$$h(r, w) - h(q, w) = g(a, w)(r - q) = a^{L(a, w)}(r - q),$$

which is zero only when $a^{L(a, w)} = 0$. This is the intersection point we have already found. A duality between points and linear complex-valued functions exists when the pairs $(z, f(z))$ are joined by the curves (also called concho spirals) specified by (11.5), as shown in Figure 11.29.

The position of the cone's apex has properties reminiscent of $\|\cdot\|$ -coords for \mathbb{R}^2 , where the horizontal position of the *point* representing a *line* reveals the slope of the

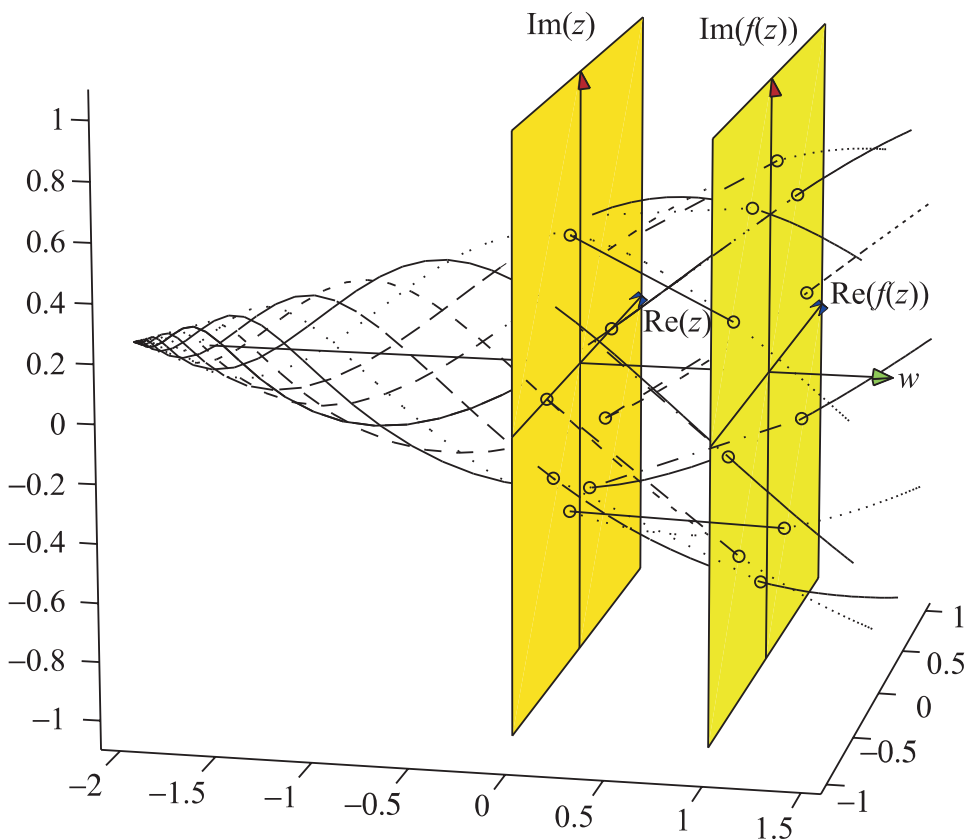


Figure 11.29. Representing the function $f(z) = (1 + i)z$ by joining the pairs $z, f(z)$ with concho spirals given by (11.5).

line. Here, the w coordinate of the cone's apex reveals the absolute value of the slope rather than the slope itself. The w axis can be divided into three sections, similar to \parallel -coords. A point whose w coordinate is less than 0 corresponds to a slope whose absolute value is larger than 1, while a w with coordinate greater than 1 corresponds to a slope with absolute value in the interval $(0, 1)$. It is clear from (11.5) that points in the interval $(0, 1)$ must be excluded; if they are not, then $0 < \frac{1}{1-|a|} < 1$ leads to a contradiction: $a < 0$.

It was pointed out earlier that duality is obtained as for \parallel -coords when the imaginary parts are zero. On the real axis the origin can be approached in only two directions. On the complex plane the origin can be reached via an infinite number of *angles*, and hence the conditions above pertain to absolute values. A more

visual explanation is that a spiral whose cone has its apex between the parallel planes (where $0 < w < 1$) only intersects *one* of the planes and therefore, it can not represent a *pair* in \mathbb{C}^2 point. For lines with slope having absolute value 1, the corresponding cone becomes a cylinder with a spiral that looks like a spring coiling from $-\infty$ to ∞ over w . To clarify further, a line in \mathbb{C}^2 is represented by a point in \mathbb{R}^3 . The *four* parts (real and imaginary) of (z_1, z_2) are mapped into *three* coordinates of the representing point in \mathbb{R}^3 . Hence we get information about $|a|$ rather a . In short, we cannot have a full point-line duality. But what we have allows us to visualize complex-valued functions.

11.5.4 Holomorphic Functions

Next we represent holomorphic (also called analytic) functions $f(z)$, that is, those whose derivative exists at every point. At each point $(z, f(z))$, the “complex tangent” is computed and then mapped into its representing point, just as for $\|\text{-coords}$. The collection of these points gives us the representing surface in \mathbb{R}^3 . The complex tangent at a point $(q, f(q))$ is described by the equation $g(z) = f'(q)z - f'(q)q + f(q)$. In this equation $a = f'(q)$ and $b = -f'(q)q + f(q)$. So by (11.6), the representing point is

$$(\Re(p), \Im(p), w),$$

where

$$p = \frac{-f'(q)q + f(q)}{1 - f'(q)}, w = \frac{1}{1 - |f'(q)|}.$$

If $f(z)$ is linear then it is its own complex tangent and therefore is represented by the same point.

Inversion

Inversion is a common mapping and our first holomorphic function visualization example. It is defined by

$$f(z) = \frac{1}{z}.$$

In polar coordinates, this is

$$f(\rho e^{i\theta}) = (\rho e^{i\theta})^{-1} = \frac{1}{\rho} e^{-i\theta},$$

showing that vector's absolute value is inverted and the angle is reflected. Actually, the following variation of inversion is far more interesting:

$$f(z) = \frac{r^2}{z}, \quad r \in \mathbb{C}. \quad (11.7)$$

In order to visualize it, p and w are computed:

$$\begin{aligned} f'(z) &= -\frac{r^2}{z^2} \\ p(q) &= \frac{-f'(q)q + f(q)}{1 - f'(q)} = \frac{\frac{r^2}{q^2}q + \frac{r^2}{q}}{1 + \frac{r^2}{q^2}} \\ &= \frac{\frac{2r^2}{q}}{\frac{q^2 + r^2}{q^2}} = \frac{2r^2 q}{q^2 + r^2} \\ w(q) &= \frac{1}{1 - |f'(q)|} = \frac{1}{1 - \left| \frac{r^2}{q^2} \right|}. \end{aligned}$$

Happily, the resulting surface reveals some properties of the inversion mapping. The symmetry between z and $f(z)$ in (11.7) is seen in Figure 11.30. It is clear that when $f(z) = \frac{r^2}{z}$, $z = \frac{r^2}{f(z)}$.

The second observation is that the points $(0, 0, 0)$ and $(0, 0, 1)$ are not part of the surface but are in the closure of it. These points represent the limit of the surface tangents as $q \rightarrow 0$ and $q \rightarrow \infty$. Third, the vertical section of the curve, which is $r \frac{2te^{i\theta}}{1+t^2e^{2i\theta}}$ for $w = \frac{1}{1-t^{-1}}$, $t \in \mathbb{R}$, consists of two nearly circular arcs. Fourth, a complex r adds a rotation factor to the inversion, causing the surface to rotate around the w -axis. The r^2 term doubles the angle.

Möbius Transformation

This also known as linear fractional or holomorphic transformation. It is defined by

$$f(z) = \frac{az + b}{cz + d}.$$

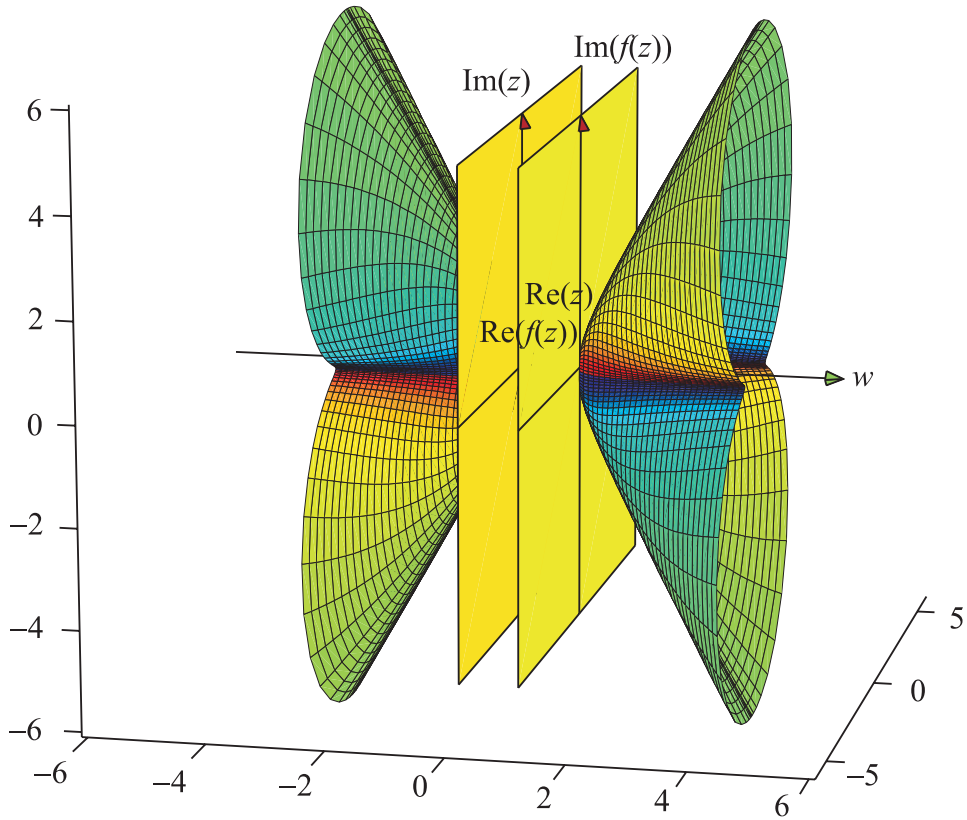


Figure 11.30. Inversion mapping $f(z) = \frac{1}{z}$.

The computation

$$\begin{aligned}
 f'(z) &= \frac{ad - bc}{(cz + d)^2} \\
 p(q) &= \frac{-\frac{ad-bc}{(cq+d)^2}q + \frac{aq+b}{cq+d}}{1 - \frac{ad-bc}{(cq+d)^2}} = \frac{bcq - adq + (aq + b)(cq + d)}{(cq + d)^2 - ad + bc} \\
 &= \frac{acq^2 + 2bcq + bd}{(cq + d)^2 - ad + bc}
 \end{aligned}$$

$$w(q) = \frac{1}{1 - \left| \frac{ad-bc}{(cq+d)^2} \right|}$$

provides the image in Figure 11.31. The ratios of the ovals' diameters to their intersecting curves have analytical significance that would take too long to explain here. Observe that inversion is a specific instance of the Möbius mapping, where $a = 0$; $b = r^2$; $c = 1$; $d = 0$.

Power Function

The power function $f(z) = z^n$, $n \in \mathbb{N}$, is yet another holomorphic mapping for which we can calculate p and w :

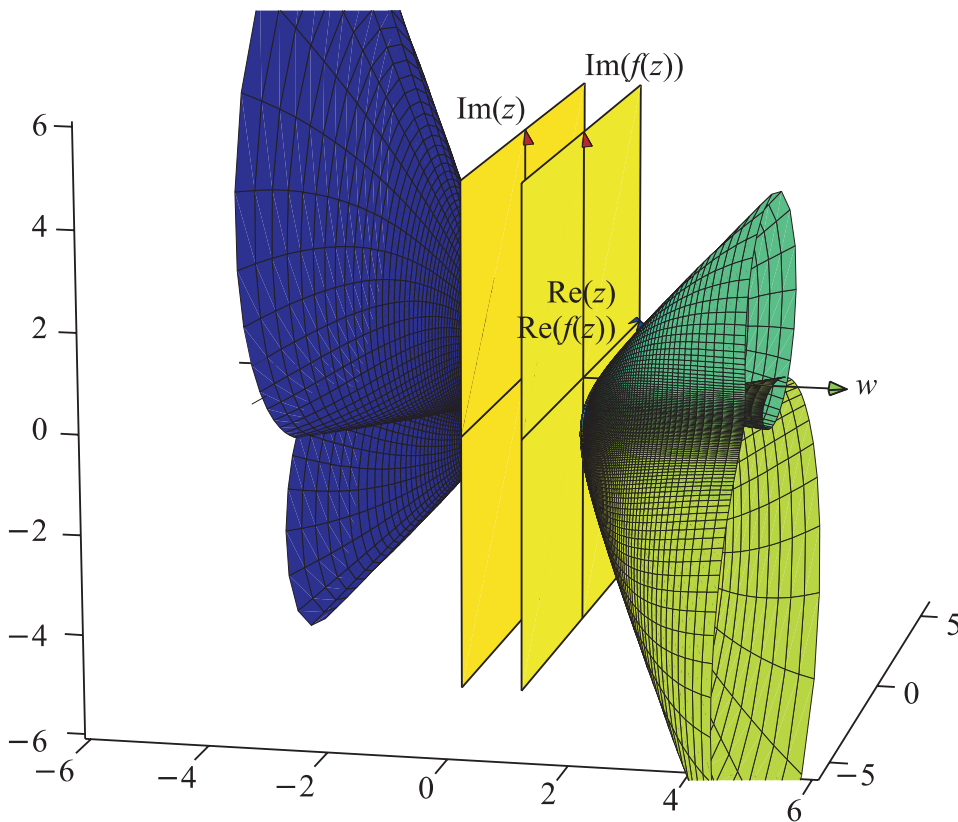


Figure 11.31. Möbius mapping $f(z) = \frac{1z+1+i}{iz+i}$.

$$f'(z) = nz^{n-1}$$

$$p(q) = \frac{-f'(q)q + f(q)}{1 - f'(q)} = \frac{-nq^{n-1}q + q^n}{1 - nq^{n-1}} = \frac{(1-n)q^n}{1 - nq^{n-1}}$$

$$w(q) = \frac{1}{1 - |f'(q)|} = \frac{1}{1 - |nq^{n-1}|} = \frac{1}{1 - n|q|^{n-1}}.$$

We consider at first the cases where $n \geq 2$. As can be seen in Figure 11.32, the surfaces are not symmetric. For $w \geq 1$ we get $n - 1$ cones that all share the same apex $(0, 0, 1)$. At this point, the derivative of the function is zero. On the opposite side, where $w \leq 0$, the number of cones is the same, but instead of a single apex they merge into an almost flat disc. This disc gets bigger as $w \rightarrow 0$, corresponding to the function values approaching infinity. The number of cones is actually $|n - 1|$, for it is determined by the derivative $f'(z)$ used to construct the picture. Hence, for inversion, where $n = -1$, there are *two* cones. Furthermore, the image of the inverse function f^{-1} can actually be obtained from the mirror image of the function, but this requires some care. Notice that in $p(q)$ there is a q^n term, and for $0 < n < 1$, there are n choices for this root. To include all these choices we need to include all z for $0 < \theta < 2n\pi$. Specifically, in Figure 11.33, $n = 1/2$ results in a “half-cone,” which is exactly half the image of its inverse z^2 .

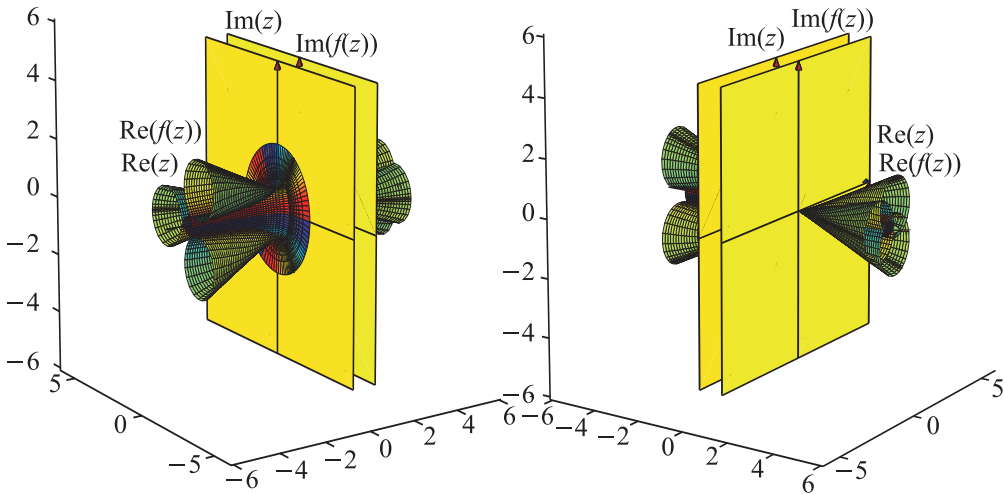


Figure 11.32. The function $f(z) = z^4$.

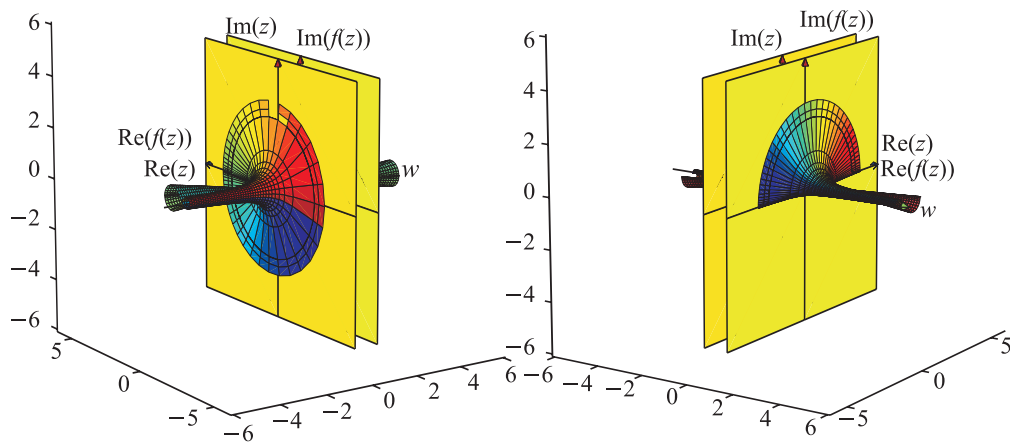


Figure 11.33. z^2 (left) versus $z^{1/2}$ (right).

Actually, this introductory portion is only indicative of a much richer collection of functions that have been studied. Visualization of \mathbb{C}^2 reveals properties on symmetries, inverses, nature of singularities and function structure in general, and this is just the beginning.

Solutions to Selected Exercises

12.1 Solutions to Exercises in Chapter 2

Exercises on Page 8

1. *Using compass only*

(a) Construction of collinear points:

Draw a circle with any center B and any radius r . Choose a point A on the circle. Draw a circle with center A of radius r . Both circles intersect at C . With C as the center draw an arc on circle B to obtain the point D . With D as the center draw an arc to obtain the point E . The points A, B, E are collinear.

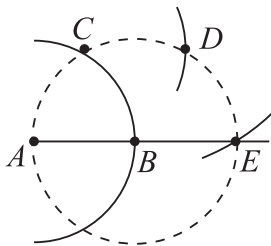


Figure 12.1. Construction of collinear points.

Proof. We shall prove that $\angle ABE = 180^\circ$:
 $\triangle ABC, \triangle BCD, \triangle BDE$ are equilateral triangles
 $\Rightarrow \angle ABC = \angle CBD = \angle DBE = 60^\circ$
 $\Rightarrow \angle ABE = 180^\circ$. ■

- (b) Given a segment \overline{AB} construct a point E such that $\overline{AB} = 2 \cdot \overline{AE}$.

It is the same construction as the previous solution. It is obvious that if $\overline{AB} = r$ then $\overline{AE} = 2r$.

- (c) *Bisecting a segment:*

Construct point C such that $AC = 2AB$ (as in Exercise 1a). Construct an arc of radius CA with center C . This will intersect the circle of radius AB centered at A at points D and E . Draw two arcs of radius DA and EA centered at D and E , respectively. Besides A , the two arcs intersect at another point F , which is the midpoint of the segment AB .

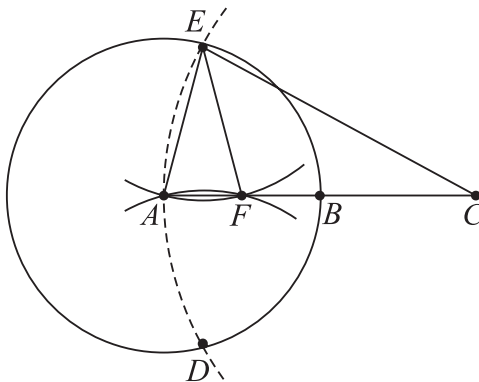


Figure 12.2. Bisecting a segment.

Proof. $\triangle AEF \sim \triangle ECA$ (\sim is “similar”); they are both isosceles with $\angle EAC$

$$\Rightarrow \frac{AC}{AE} = \frac{AE}{AF}. \text{ We also know } AC = 2 \cdot AB, AE = AB \Rightarrow AC = 2 \cdot AE$$

$$\Rightarrow \frac{AC}{AE} = 2 \Rightarrow AB = AE = 2 \cdot AF. \quad \blacksquare$$

- (d) *Trisecting a segment:*

Construct point D such that $AD = 3 \cdot AB$. Construct an arc of radius DA with center D . It will intersect the circle of radius AB centered at A at points M and E . Draw two arcs of radius MA and EA centered at M and E , respectively. Besides A , the two arcs will intersect at another point F , which is $\frac{AB}{3}$.

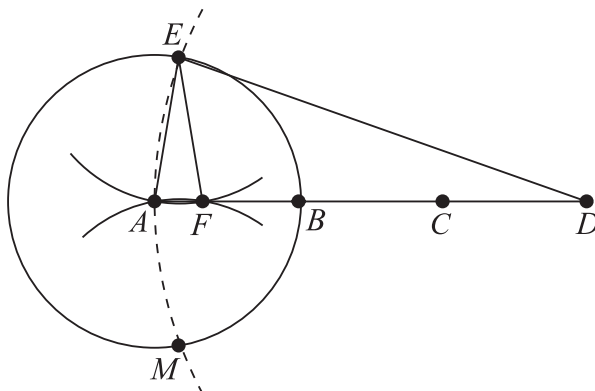


Figure 12.3. Trisecting a segment.

Proof. $\triangle AEF \sim \triangle EDA$; they are both isosceles with $\angle EAD$

$$\Rightarrow 3 = \frac{3AB}{AB} = \frac{AD}{FE} = \frac{AE}{FA} = \frac{AB}{FA}$$

$$\Rightarrow AB = 3 \cdot AF.$$



- (e) Constructing the point of intersection between circle C and line l (given by two points A and B):

With centers A and B and radius AO and BO , respectively, draw two arcs.

Case 1. If the two arcs intersect *only* at point O , the center O lies on the line AB . The arc centered at A intersects the circle at P and Q , and the arc centered at B intersects the circle at M and N . Now we have to find the midpoints of those two arcs PQ and MN on the circle.

Let us find the midpoint of PQ with center O :

From P and Q as centers draw two arcs of radius PO . From O draw arcs OT and OK equal to PQ , then draw two arcs with TP and KQ as radii and with T and K as centers intersecting at R . Finally, with OR as radius draw an arc from T ; it intersects the arc PQ at the midpoint X , which is also the intersection point of the circle and the line AB . The same construction has to be done on the arc MN in order to find the other intersection point X' .

Case 2. If those two arcs intersect at O and at another point H , draw another circle with center H with the same radius of circle O . Both circles intersect at X and X' as required.

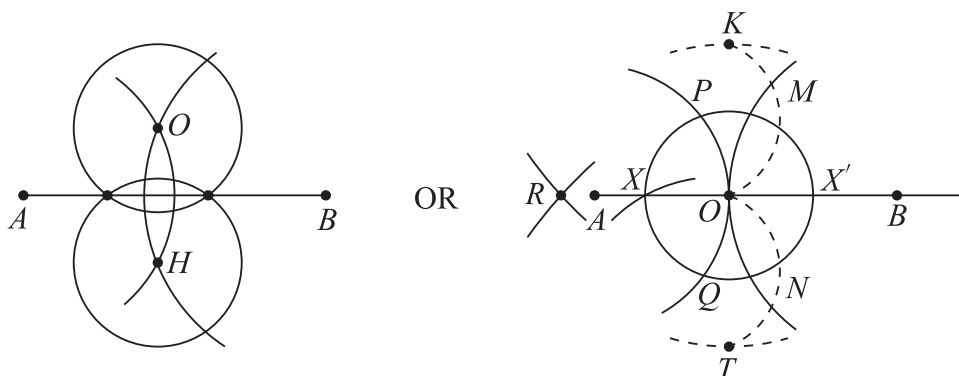


Figure 12.4. Intersection between a circle and a line.

Remark. Finding the *inverse* point of a given point A with respect to a given circle C with center O and radius r :

Case 1. $OA > \frac{1}{2} \cdot r$.

Draw an arc with center A of radius AO ; the arc intersects the circle at two points C and D . Draw two arcs centered at C and D of radius CO ; they meet again at A' , which is the inverse point of A .

Case 2. $OA \leq \frac{1}{2} \cdot r$.

Find the point A_1 such that $OA_1 = 2 \cdot OA$ (as in Exercise 1a).

If $OA_1 > \frac{1}{2} \cdot r$ find the inverse A'_1 of A_1 by *Case 1*. If $OA_1 \leq \frac{1}{2} \cdot r$, we must go on doubling, finding in succession A_2, A_3, \dots, A_p in OA , where $OA_2 = 2 \cdot OA_1, OA_3 = 2 \cdot OA_2, \dots, OA_p = 2 \cdot OA_{p-1} = 2^p \cdot OA$ until $OA_p > \frac{1}{2} \cdot r$. We then find the inverse A'_p of A_p , and again double p times, constructing in succession the points $A'_{p-1}, \dots, A'_2, A'_1, A'$, where $OA' = 2 \cdot OA'_1 = 2^p \cdot OA'_p$.

Then A' is the inverse of A .

Remark. Finding the *inverse* circle of a given circle C with respect to a segment AB :

Draw two arcs centered at A and B with radii AO and BO , respectively. They intersect again at point H .

From H as the center draw an arc with radius HO that intersects the circle C at points M and N .

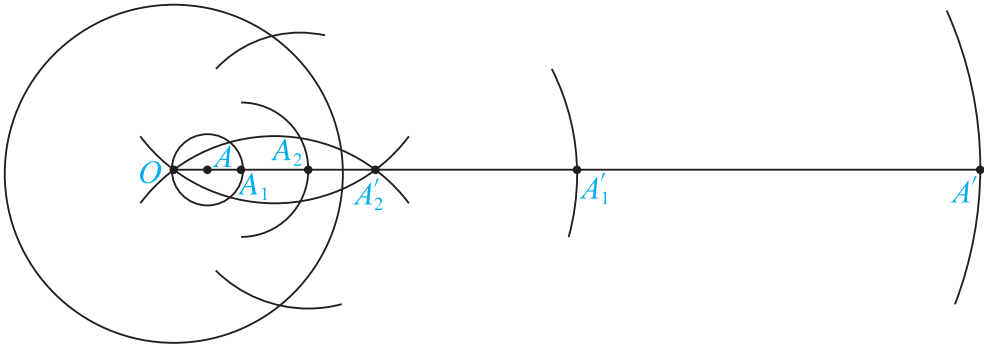


Figure 12.5. Inverse point with respect to the circle.

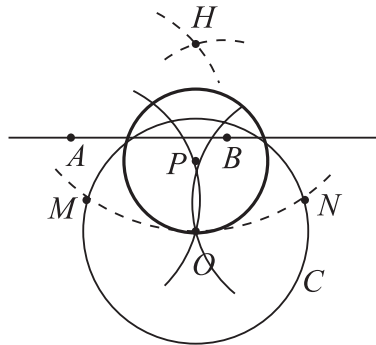


Figure 12.6. Inverse circle with respect to segment AB .

From M and N as centers draw two arcs with radius MO ; they again intersect at point P . The circle with center P and with radius OP is the *inverse* circle of C with respect to AB . (Figure 12.6)

- (f) Let the lines be given by AB and $A'B'$. Draw any circle C and by the preceding method find the *inverse* circles (with centers T_1, T_2) of C with respect to AB and $A'B'$. These circles intersect at O and Y . The point X *inverse* to Y is the required point of intersection, and can be constructed by the process explained before. That X is the required point is evident from the fact that Y is the only point that is inverse to a point of both AB and $A'B'$. Hence the point X inverse to Y must lie on both AB and $A'B'$.

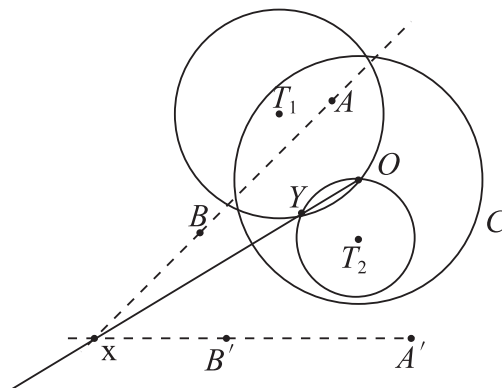


Figure 12.7. Intersection between two lines.

2. *Using a single circle and a straightedge*

Given an angle A , provide a construction for finding an angle $= \frac{1}{2} A$:

Put the center of the circle at point A ; call the intersection points between them B and D . Continue AB until it intersects the circle at C . Then $\angle BCD = \frac{\angle BAD}{2}$.

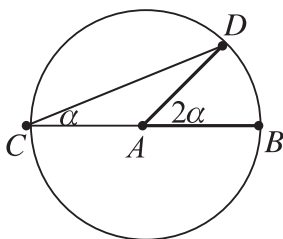


Figure 12.8. Half-angle.

3. Transform a rectangle to a square of *equal* area:

We have rectangle $ABCD$; $AB = b$, $AD = a \Rightarrow AC = DB = \sqrt{a^2 + b^2}$.

Draw a circle centered at B of radius a . Continue AB until it intersects the circle at point E ; then $EA = a + b$. Bisect the segment EA and call the midpoint F (Exercise 1c). Draw a circle with center F of radius FE . Draw an arc centered at A of radius AC intersecting the circle at M . Then $\triangle EAM$ is right-angled $\Rightarrow EM = \sqrt{2ab}$.

Now we have to construct a square with diagonal EM :

Draw two circles centered at M and E of radius EM ; they intersect at V and U . Using the ruler construct the segment UV ; it intersects segment EM on W .

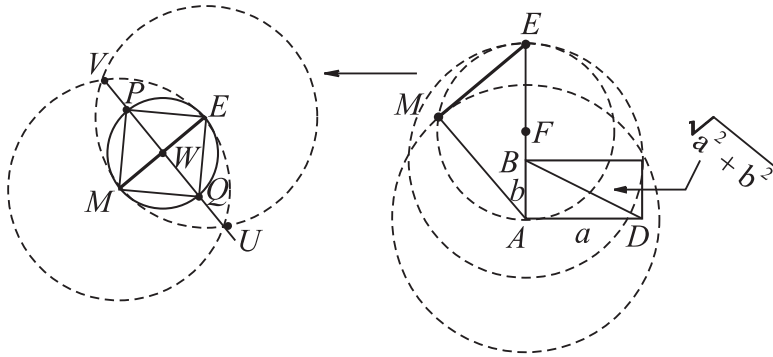


Figure 12.9. The rectangle and the square have the same area.

Draw a circle with center W of radius WE ; it intersects segment UV at P and Q . We have a square $EQMP$ of area $= 2 \cdot ab$.

4. Given a square with area A , construct a square with area $2 \cdot A$:

We know that the diagonal of the square $FBCD$ is $\sqrt{2A}$. Draw a circle at center C of radius DC intersecting points M and N . The square $BDMN$ has area $2 \cdot A$.

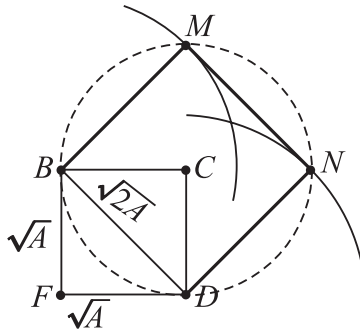


Figure 12.10. Doubling the area of a square.

5. To prove that *a triangle having two equal angle bisectors is isosceles*, it suffices to prove the following theorem.

Theorem. *In a triangle having two different angles, the smaller angle has the longer bisector.*

Proof. In $\triangle ABC$, AE and BD are the angle bisectors with angles $\angle B > \angle A$. Draw a circle through the three points A , D , and B that intersects AE at a point F . The angles $\angle DAF$ and $\angle DBF$ must be equal (i.e., $\frac{1}{2} \cdot A$), since they subtend

the same arc DF . By construction, angles $\angle DAB < \angle FBA \Rightarrow$ that the arcs $DB < AF \Rightarrow$ the chords $DB < AF < AE$. ■

The above result is sometimes called the “Steiner–Lehmus theorem.” The problem was first posed in 1840 by Lehmus to Sturm, who passed it on to the famous geometer Steiner, who apparently gave the first satisfactory proof. Papers appeared in various journals in 1842, 1844, 1848, 1854–1864, and many others since. Henderson wrote surveys in *Scripta Mathematica*, Vol. 21, 1955, 223–232, 309–312, and there are others. The proofs given are indirect, i.e., based on the *contrapositive*, which has upset some people, though Euclid himself used this method of proof extensively. There exist several (60+) different proofs of this result. The above, after some modifications, is mine. Enjoy!

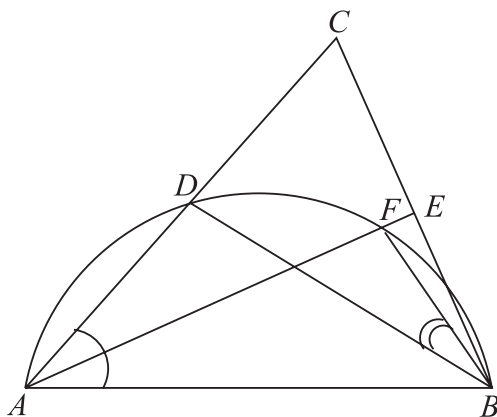


Figure 12.11. “Steiner–Lehmus theorem.”

Exercises on Page 13

1. The line l and point C determine a plane π_l in \mathbb{R}^3 that intersects the hemisphere in a great circle, c_ℓ . This semicircle is the image of the line.
2. Let θ be the angle between two lines, l_1 , and l_2 , on the plane. The angle between planes π_{l_1} and π_{l_2} is also θ , and therefore, so is the angle between the semicircles c_{ℓ_1} and c_{ℓ_2} , which are the images of the two lines on the hemisphere.
3. Circles are mapped to circles. An intersection of a cone with a hemisphere is a circle.
4. The upper line is called l_1 and the other is l_2 . Using the ruler draw any other line l_3 . Choose two points, point A on line l_1 , point K on line l_3 . Draw the segment AK . The intersection point with l_2 is C . Choose another point A' on l_1 . Draw KA' . The intersection with l_2 is C' . Draw $A'R$ and continue until it intersects

the line l_3 at M . Continue RC' until intersection with l_3 at Q . QC intersects AM at R' . On the line RR' there is also the point T . (Desargues's theorem)

5. The same as before, but the lines should be very close so that all the points can be drawn with our short ruler. Note to student: Try to give a better construction than this!

Exercises on Page 16

1. Ideal points satisfy the axioms of points (Axioms 1 + 2):
 - Two distinct ideal points are on a unique ideal line.
 - Three distinct noncollinear ideal points are in a unique ideal plane.

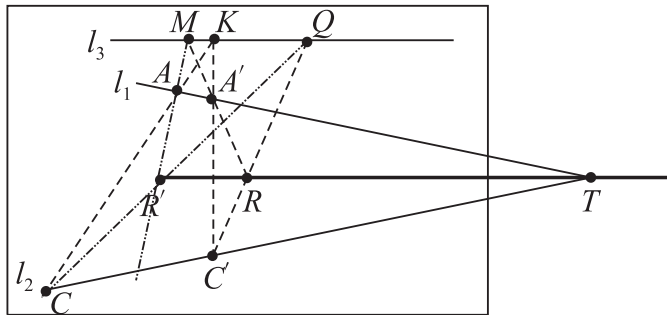


Figure 12.12. Exercise 4.

2. Augmented lines satisfy the axioms of lines (Axioms 3 + 4):
 - Let $(l, L_\infty), (p, P_\infty)$ be two coplanar augmented lines. If $l \parallel p$ they have the same direction \Rightarrow the same ideal point $\Rightarrow L_\infty = P_\infty$. If $l \not\parallel p$ they intersect at an ordinary unique point.
 - Let $(l, L_\infty), (\pi, p_\infty)$ augmented line and augmented plane such that $l \not\subset \pi$. If $l \parallel \pi \Rightarrow L_\infty \in p_\infty \Rightarrow (l, L_\infty) \cap (\pi, p_\infty) = L_\infty$. If $l \not\parallel \pi$ they intersect at a unique point.
3. Augmented planes satisfy the axioms of planes (Axioms 4 + 5):
 - Let $(\pi_1, p_\infty) \neq (\pi_2, l_\infty)$. If $\pi_1 \parallel \pi_2 \Rightarrow$ for every line on π_1 there is a parallel line on $\pi_2 \Rightarrow \pi_1$ and π_2 have the same directions $\Rightarrow (\pi_1, p_\infty) \cap (\pi_2, l_\infty) = p_\infty$. If $\pi_1 \not\parallel \pi_2$ they intersect at a unique line, as known.
 - The line-on-plane relation is already proved above.

4. Ideal plane π_∞ satisfies the axioms of a plane:

For this we have to add two more axioms:

- (a) two distinct spaces are on a unique plane.
 (b) a space and a plane are on a unique line.

- Let (R, π_∞) be a space with ideal plane and $(\alpha, l_\infty) \not\subset (R, \pi_\infty)$.
 $(R, \pi_\infty) \cap (\alpha, l_\infty) = (p, A_\infty)$ by axiom (b) $\Rightarrow \pi_\infty \cap l_\infty = A_\infty$.
- Let $(\pi_\infty \neq \alpha_\infty)$ be two distinct ideal planes \Rightarrow they are associated to two distinct spaces (not parallel). By Axiom (a) they intersect at a unique plane, and this plane has a unique ideal line $l_\infty = \pi_\infty \cap \alpha_\infty$.

Exercises on Page 21

A *proof* for the duals using only the projective 3-space axioms:

1. Two *planes* are on a line; if they are distinct they are on a unique line according to Axiom 5.
 The dual is: Two *points* are on a line (Axiom 1).
2. Axiom 4: A *plane* π and a line ℓ not on π are on a *point*. Its dual is: A *point* P and a line ℓ are on a *plane*; choose two distinct points P_1, P_2 on ℓ ; P, P_1, P_2 are noncollinear; according to Axiom 2, they are in a plane.
3. The sets of points and lines satisfy the projective plane axioms.

Proof.

1. Any two distinct points A, B on a unique line satisfy Axiom 1.
2. If p, q are distinct lines, there is at least one point on both; there is one point on both; this satisfies Axiom 3.
3. There are at least three distinct points on any line \Rightarrow there are two distinct points on a line; satisfies Axiom 1.
4. Not all points are on the same line. This means that every three noncollinear points are on a plane; satisfies Axiom 2. ■
4. Let P be any point. By Axiom 1, every point on π determines a line together with P , which obviously passes through P . Hence all the points of π lie on the n lines incident with P . Each of these lines is incident with $n - 1$ points other than P , so there are $n(n - 1)$ of them, and including P , we get $n^2 - n + 1$ points in π .
5. *The algorithm:*
Step ($k = 1$): Construct a matrix $M_{n \times (n-1)}$: the first row R_1 contains the numbers $2, \dots, n$, every other row R_i contains the next $(n - 1)$ numbers. The first n lines connect the number 1 with the numbers of $R_i (i = 1, \dots, n)$.

Step ($k = 2$): Construct a matrix $A_{(n-1) \times (n-1)}$: $A = [M \setminus R_1]^T$. Define the columns of A : C_1, \dots, C_{n-1} . The next $(n - 1)$ lines connect the number 2 with every row of A .

Step ($k \geq 3$): Define A' : the numbers on the diagonal of A will be the first row of A' . For each of the remaining C_i , the columns are permuted by one element in cyclic order. The next $(n - 1)$ lines will connect k with the rows of A' .

Do this step until $k = n$. The process is clarified with the example below.

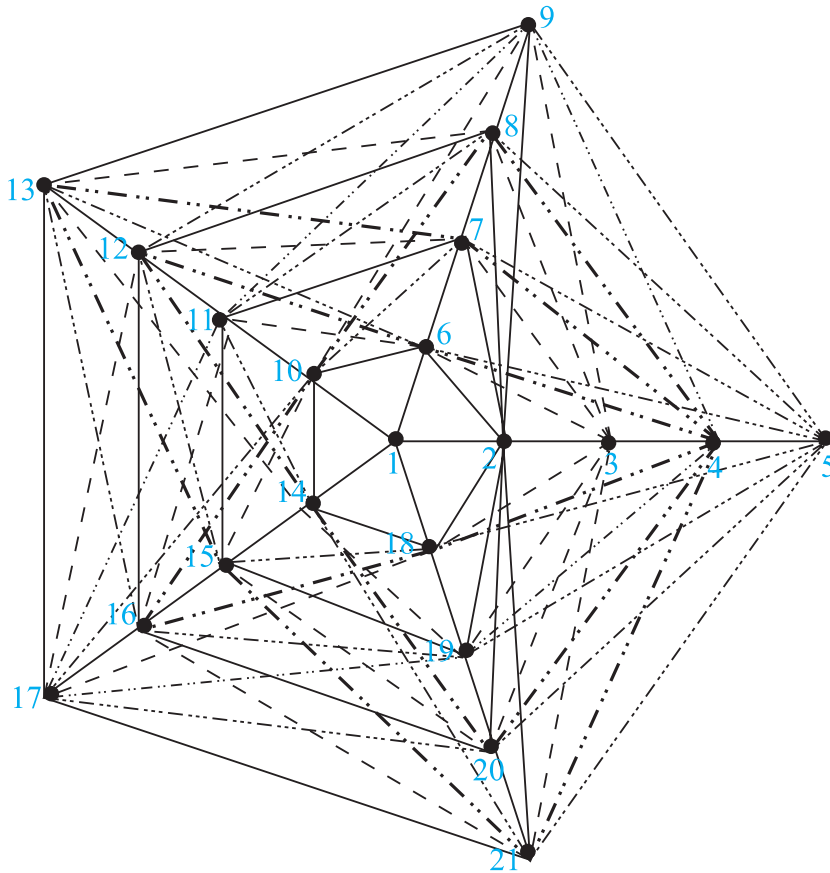


Figure 12.13. Finite geometry of five points.

6. *The 5-point Geometry (21 points):*

	2	3	4	5		6	10	14	18		6	11	16	21
	6	7	8	9		7	11	15	19		7	12	17	18
1 \longrightarrow	10	11	12	13	2 \longrightarrow	8	12	16	20	3 \longrightarrow	8	13	14	19
	14	15	16	17		9	13	17	21		9	10	15	20
	18	19	20	21										
	6	12	14	20		6	13	16	19					
	7	13	15	21		7	10	17	20					
4 \longrightarrow	8	10	16	18	5 \longrightarrow	8	11	14	21					
	9	11	17	19		9	12	15	18					

The 6-Point Geometry (31 points):

	2	3	4	5	6		7	12	17	22	27		7	13	19	25	31
	7	8	9	10	11		8	13	18	23	28		8	14	20	26	27
1 \longrightarrow	12	13	14	15	16	2 \longrightarrow	9	14	19	24	29	3 \longrightarrow	9	15	21	22	28
	17	18	19	20	21		10	15	20	25	30		10	16	17	23	29
	22	23	24	25	26		11	16	21	26	31		11	12	18	24	30
	27	28	29	30	31												
	7	14	21	23	30		7	15	18	26	29		7	16	20	24	28
	8	15	17	24	31		8	16	19	22	30		8	12	21	25	29
4 \longrightarrow	9	16	18	25	27	5 \longrightarrow	9	12	20	23	31	6 \longrightarrow	9	13	17	26	30
	10	12	19	26	28		10	13	21	24	27		10	14	18	22	31
	11	13	20	22	29		11	14	17	25	28		11	15	19	23	27

Exercises on Page 29

1. Create 24 distinct permutations of four symbols A, B, C, D . According to the cross ratio,

$$(A, B; C, D) = (B, A; D, C) = (C, D; A, B) = (D, C; B, A) = \frac{AC}{BC} \cdot \frac{BD}{AD},$$

$$(A, B; D, C) = (B, A; C, D) = (D, C; A, B) = (C, D; B, A) = \frac{AD}{BD} \cdot \frac{BC}{AC},$$

$$(A, C; B, D) = (C, A; D, B) = (B, D; A, C) = (D, B; C, A) = \frac{AB}{CB} \cdot \frac{CD}{AD},$$

$$(C, B; A, D) = (B, C; D, A) = (A, D; C, B) = (D, A; B, C) = \frac{CA}{BA} \cdot \frac{BD}{CD},$$

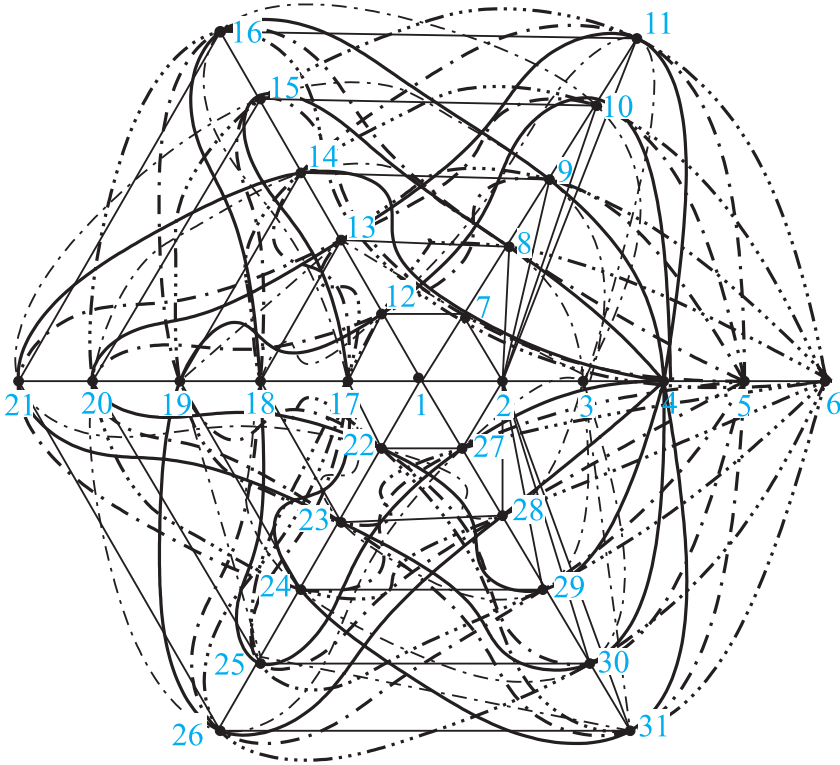


Figure 12.14. Finite geometry of six points.

$$(D, B; A, C) = (B, D; C, A) = (A, C; D, B) = (C, A; B, D) = \frac{DA}{BA} \cdot \frac{BC}{DC},$$

$$(B, C; A, D) = (C, B; D, A) = (A, D; B, C) = (D, A; C, B) = \frac{BA}{CA} \cdot \frac{CD}{BD}$$

\Rightarrow There are six sets of permutations having the same cross ratio values. ■

2. The cross ratio is invariant under a perspectivity with respect to an ideal point

P_∞ :

Given two lines

$$l_1(A, B, C, D) \stackrel{P_\infty}{\underset{\wedge}{\sim}} l_2(A', B', C', D'),$$

P_∞ is an ideal point $\Rightarrow AA' \parallel BB' \parallel CC' \parallel DD'$.

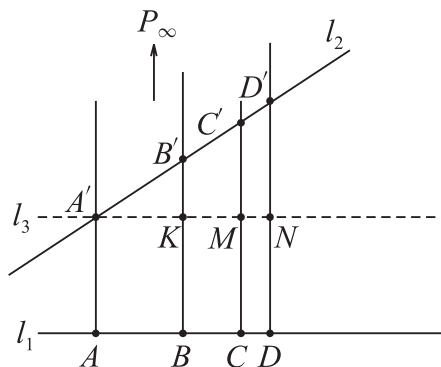


Figure 12.15. Exercise 2.

Construct a line $l_3 \parallel l_1$; then $AB = A'K$, $BC = KM$, $CD = MN$

$$\Rightarrow \triangle A'B'K \sim \triangle A'C'M$$

$$\Rightarrow \frac{A'B'}{A'C'} = \frac{AB}{AC}$$

$$\Rightarrow \frac{A'C' - B'C'}{A'C'} = \frac{AC - BC}{AC}$$

$$\Rightarrow 1 - \frac{B'C'}{A'C'} = 1 - \frac{BC}{AC}$$

$$\Rightarrow \frac{B'C'}{A'C'} = \frac{BC}{AC}.$$

In the same way, $\frac{BD}{AD} = \frac{B'D'}{A'D'}$, so the cross ratio is invariant:

$$\Rightarrow (A, B; C, D) = (A', B'; C', D').$$

3. **Theorem 2.** *A projectivity is determined by the correspondence between two triples of distinct points.*

Lemma 1. *It is possible, by a sequence of not more than three perspectivities, to relate any three distinct collinear points to any other three distinct collinear points.*

Proof of Lemma 1. If the two triads ABC and $A'B'C'$ are on distinct lines, let R, S, C_0 denote the points where the respective lines AA', BB', BA' meet CC' . Then ABC and $A'B'C'$ are related by the sequence of two perspectivities

$$ABC \stackrel{R}{\overline{\wedge}} A'BC_0 \stackrel{S}{\overline{\wedge}} A'B'C'.$$

(If A and A' coincide, we merely use the perspectivity from S .)

If the two triads are on one line, we use a quite arbitrary perspectivity $ABC \stackrel{\infty}{\overline{\wedge}} A_1B_1C_1$ to obtain a triad on another line and then relate $A_1B_1C_1$ to $A'B'C'$ by the above construction. ■

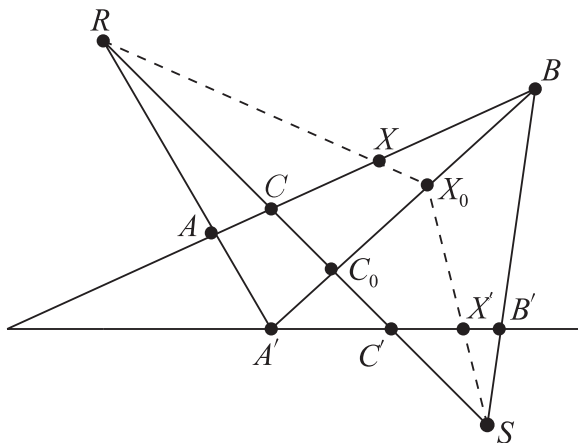


Figure 12.16. Lemma 1.

Lemma 2. *A projectivity having more than two invariant points can only be the identity.*⁵⁸

Proof of Theorem 2. Suppose we are given three points A, B, C of one range and corresponding points A', B', C' of the other. We proceed, as in the proof of Lemma 1, to construct a projectivity $X \overline{\wedge} X'$ such that $ABC \overline{\wedge} A'B'C'$.

To establish the uniqueness of this projectivity, we have to prove that a different construction (e.g., by joining AB' instead of BA') would yield the same X' for a given X . Suppose one construction gives $ABCX \overline{\wedge} A'B'C'X'$, while another gives $ABCX \overline{\wedge} A'B'C'X'_1$. Then by combining the two constructions we obtain

$$A'B'C'X' \overline{\wedge} A'B'C'X'_1.$$

This combined projectivity has three invariant points A', B', C' .

Hence by Lemma 2, X'_1 must coincide with X' . ■

⁵⁸H.S.M. Coxeter. *The Real Projective Plane*, pp. 40–42.

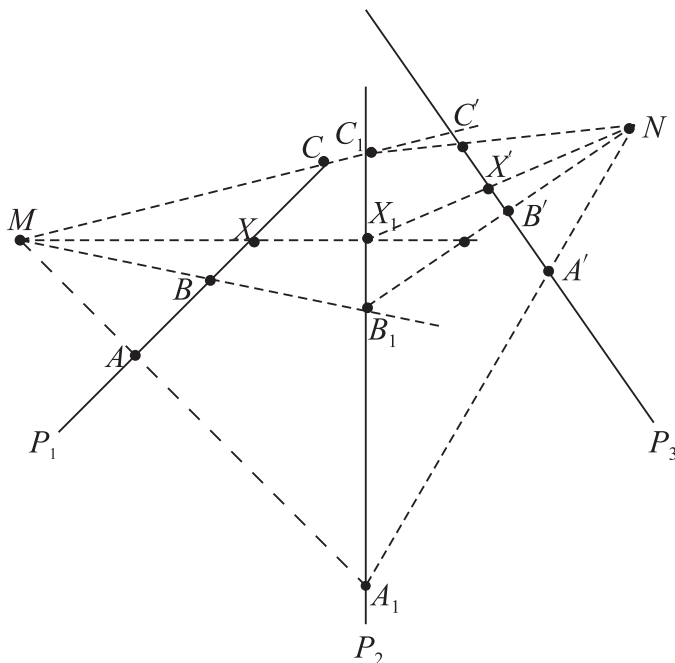


Figure 12.17. Theorem 2.

Lemma 3. Let $l \overset{O}{\underset{\wedge}{\parallel}} m \overset{P}{\underset{\wedge}{\parallel}} n$, with $l \neq n$. Then there are a line m' and points $O' \in n$, $P' \in l$ such that $l \overset{O'}{\underset{\wedge}{\parallel}} m' \overset{P'}{\underset{\wedge}{\parallel}} n$ gives the projectivity $l \underset{\wedge}{\parallel} n$.

Proof of Lemma 3. Let l, m, n, O, P be given. Let A, A' be two points on l , and let

$$AA' \overset{O}{\underset{\wedge}{\parallel}} BB' \overset{P}{\underset{\wedge}{\parallel}} CC'.$$

Let OP intersect n at O' . Since we assumed $O, P, l \cap n = X$ are not collinear, $O' \neq X$, so $O' \notin l$. Draw $O'A, O'A'$, and let them intersect PC, PC' at D, D' , respectively. Now corresponding sides of the triangles ABD and $A'B'D'$ intersect at O, P, O' , respectively, which are collinear; hence the lines joining corresponding vertices are concurrent. Thus m_1 passes through the point $Y = l \cap m$.

Thus m_1 is determined by D and Y , so as A' varies, D' varies along the line m_1 . Thus our original projectivity is equal to the projectivity

$$l \overset{O'}{\underset{\wedge}{\parallel}} m_1 \overset{P}{\underset{\wedge}{\parallel}} n.$$

Performing the same argument again, we can move P to $P' = OP \cap l$, and find a new line m' such that

$$l \stackrel{O'}{\underset{\wedge}{\sim}} m' \stackrel{P'}{\underset{\wedge}{\sim}} n$$

gives the original projectivity. ■

4. **Theorem 3.** *A projectivity can be expressed by at most two perspectivities.*

Proof. A projectivity was defined as a composition of an arbitrary chain of perspectivities. Thus it will be sufficient to show, by induction, that a chain of length $n > 2$ can be reduced to a chain of length $n - 1$, i.e., it will be sufficient to prove that a chain of three perspectivities can be reduced to a composition of two perspectivities.

The argument of Lemma 3 actually shows that the line m can be moved so as to avoid any given point. Thus one can see easily that it is sufficient to prove the following:

Let

$$\ell \stackrel{P}{\underset{\wedge}{\sim}} m \stackrel{Q}{\underset{\wedge}{\sim}} n \stackrel{R}{\underset{\wedge}{\sim}} o$$

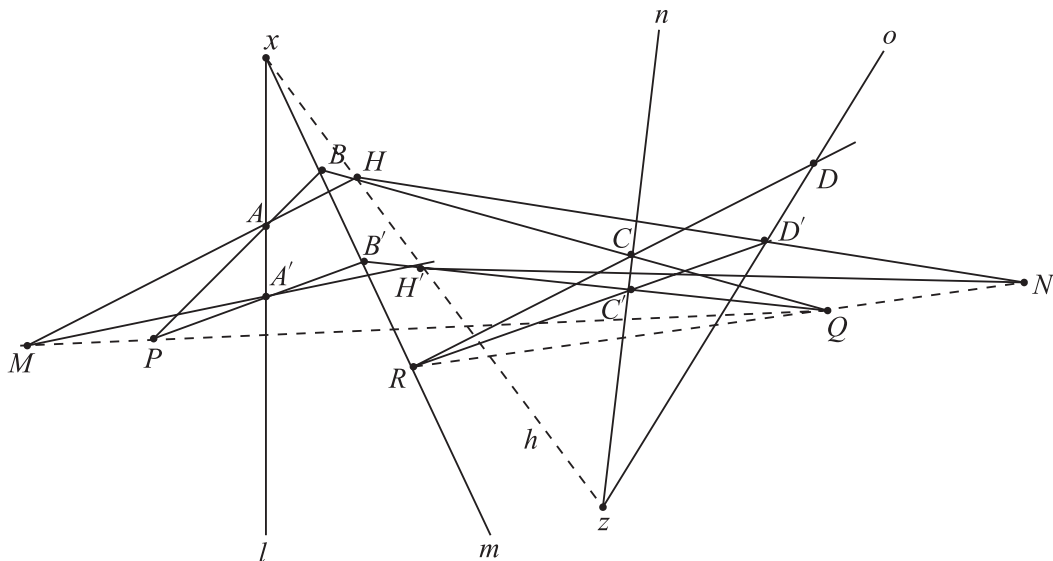
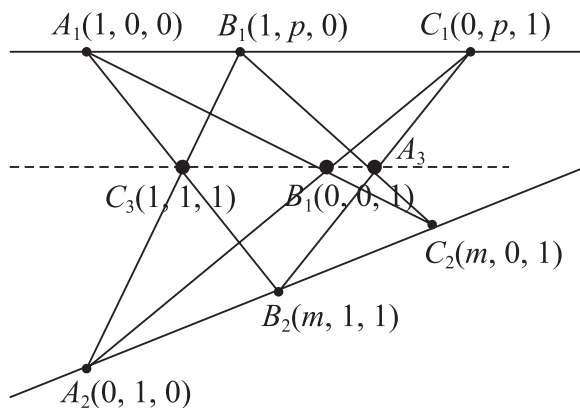
be a chain of three perspectivities, with $\ell \neq o$. Then the resulting projectivity $\ell \underset{\wedge}{\sim} o$ can be expressed as a product of at most two perspectivities.

If two of the four lines ℓ, m, n, o coincide, then there are only two perspectives. Second, we have either $m \underset{\wedge}{\sim} o$, in which case we are done, or n can be moved (Lemma 3) so that the centers of the perspectivities $m \underset{\wedge}{\sim} n$ and $n \underset{\wedge}{\sim} o$ are at o, m respectively. So we have

$$\ell \stackrel{P}{\underset{\wedge}{\sim}} m \stackrel{Q}{\underset{\wedge}{\sim}} n \stackrel{R}{\underset{\wedge}{\sim}} o$$

with ℓ, m, n, o all distinct, $Q \in o$, and $R \in m$. Let $X = \ell \cap m$, $Z = n \cap o$, and draw $h = XZ$. We may assume that $X \notin o$ (indeed, we could have moved m , by Lemma 3 to make $X \notin o$). Therefore $Q \notin h$. Project $m \stackrel{Q}{\underset{\wedge}{\sim}} h$, and let $BB' \stackrel{Q}{\underset{\wedge}{\sim}} HH'$.

Now $CDH \stackrel{Z}{\underset{\wedge}{\sim}} C'D'H'$. Corresponding sides intersect at Q, R ; hence by Desargues's theorem, the remaining corresponding sides intersect at a point N on QR . Thus N is determined by DH alone, and $h \stackrel{N}{\underset{\wedge}{\sim}} o$.

**Figure 12.18.** Theorem 3.**Figure 12.19.** Pappus's theorem.

Similarly, $ABH \stackrel{X}{\underset{\wedge}{\sim}} A'B'H'$, so using Desargues's theorem again, we find that $AH \cap A'H' = M \in PQ$. Hence, $\ell \stackrel{M}{\underset{\wedge}{\sim}} h$. So we have the original projectivity represented as a composition of two perspectivities

$$\ell \stackrel{M}{\underset{\wedge}{\sim}} h \stackrel{N}{\underset{\wedge}{\sim}} o.$$

■

Exercises on Page 43

- 1.
- Analytic proof for Pappus's theorem:*

Assign homogeneous coordinates as follows:

$$A_1 = (1, 0, 0), A_2 = (0, 1, 0), C_3 = (1, 1, 1), B_3 = (0, 0, 1).$$

On line A_2C_3 choose point $B_1 = (1, p, 1)$; on line A_1C_3 choose point $B_2 = (m, 1, 1)$.

$$C_1 = A_2B_3 \cap A_1B_1 = (0, p, 1),$$

$$C_2 = A_1B_3 \cap A_2B_2 = (m, 0, 1),$$

$$B_1C_2 = \left[\begin{vmatrix} 1 & p \\ m & 0 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ m & 1 \end{vmatrix}, \begin{vmatrix} p & 1 \\ 0 & 1 \end{vmatrix} \right] = [-pm, 1 - m, p],$$

$$B_2C_1 = \left[\begin{vmatrix} 0 & p \\ m & 1 \end{vmatrix}, \begin{vmatrix} 0 & 1 \\ m & 1 \end{vmatrix}, \begin{vmatrix} p & 1 \\ 1 & 1 \end{vmatrix} \right] = [-pm, -m, p - 1],$$

$$A_3 = B_2C_1 \cap B_1C_2$$

$$= \left(\begin{vmatrix} -pm & 1 - m \\ -pm & -m \end{vmatrix}, \begin{vmatrix} -pm & p \\ -pm & p - 1 \end{vmatrix}, \begin{vmatrix} 1 - m & p \\ -m & p - 1 \end{vmatrix} \right),$$

$$A_3 = (pm, pm, p + m - 1).$$

Checking the determinant we find that $\begin{vmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ pm & pm & p+m-1 \end{vmatrix} = 0$

\Rightarrow The points A_3, B_3, C_3 are collinear. ■

- 2.
- Ideal plane of (R^3, p_∞)*

We already know that every ideal point is represented in homogenous coordinates as $(P_{1i}, P_{2i}, P_{3i}, 0)$. A plane $\pi \in (R^3, p_\infty)$ is $C_1X_1 + C_2X_2 + C_3X_3 = C_0$. In homogenous coordinates we can write it as $C_1X_1 + C_2X_2 + C_3X_3 + C_4X_4 = 0$.

Three ideal points are collinear if the determinant = 0:

$$\begin{vmatrix} X_1 & X_2 & X_3 & X_4 \\ P_{11} & P_{21} & P_{31} & 0 \\ P_{12} & P_{22} & P_{32} & 0 \\ P_{13} & P_{23} & P_{33} & 0 \end{vmatrix} = 0$$

$$\Rightarrow C \cdot X_4 = 0 \iff X_4 = 0.$$

Hence, an ideal plane is represented as $[C_1, C_2, C_3, 0]$.

Exercise on Page 46

The set of all linear fractional transformations

$$T(x) = \frac{ax + b}{cx + d}$$

with the composition operation form a group.

Proof. Closure property:

$$\begin{aligned} T(x) \circ T(y) &= \frac{a\left(\frac{a_1y+b_1}{c_1y+d_1}\right) + b}{c\left(\frac{a_1y+b_1}{c_1y+d_1}\right) + d} \\ &= \frac{(aa_1 + bc_1)y + (ab_1 + bd_1)}{(ca_1 + dc_1)y + (cb_1 + dd_1)} = \frac{My + N}{Vy + U} \in G. \end{aligned}$$

Associativity:

$$\begin{aligned} (T(x) \circ T(y)) \circ T(z) &= \left(\frac{(aa_1 + bc_1)y + (ab_1 + bd_1)}{(ca_1 + dc_1)y + (cb_1 + dd_1)} \right) \circ T(z) \\ &= \frac{(aa_1 + bc_1)\left(\frac{a_2z+b_2}{c_2z+d_2}\right) + (ab_1 + bd_1)}{(ca_1 + dc_1)\left(\frac{a_2z+b_2}{c_2z+d_2}\right) + (cb_1 + dd_1)} \\ &= \frac{(aa_1a_2 + bc_1a_2 + ab_1c_2 + bd_1c_2)z + aa_1b_2 + bb_2c_1 + ab_1d_2 + bd_1d_2}{(ca_2a_2 + dc_1a_2 + cc_2b_1 + dd_1c_2)z + ca_1b_2 + dc_1b_2 + cb_1d_2 + dd_1d_2}, \end{aligned}$$

$$\begin{aligned} T(x) \circ (T(y) \circ T(z)) &= T(x) \circ \left(\frac{a_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right) + b_1}{c_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right) + d_1} \right) = \frac{a\left(\frac{(a_1a_2+b_1c_2)z+a_1b_2+b_1d_2}{(c_1a_2+d_1c_2)z+c_1b_2+d_1d_2}\right) + b}{c\left(\frac{(a_1a_2+b_1c_2)z+a_1b_2+b_1d_2}{(c_1a_2+d_1c_2)z+c_1b_2+d_1d_2}\right) + d} \\ &= \frac{(aa_1a_2 + bc_1a_2 + ab_1c_2 + bd_1c_2)z + aa_1b_2 + bb_2c_1 + ab_1d_2 + bd_1d_2}{(ca_2a_2 + dc_1a_2 + cc_2b_1 + dd_1c_2)z + ca_1b_2 + dc_1b_2 + cb_1d_2 + dd_1d_2} \\ &\Rightarrow (T(x) \circ T(y)) \circ T(z) = T(x) \circ (T(y) \circ T(z)). \end{aligned}$$

Identity:

$$\exists e = \frac{1x + 1}{1x + 1} \in G : \forall a \in G \quad e \circ a = a \circ e = a.$$

Inverse:

$$\forall a = \frac{a_1x + b_1}{c_1x + d_1} \in G \exists a^{-1} = \frac{c_1x + d_1}{a_1x + b_1} \in G : a \circ a^{-1} = a^{-1} \circ a = e$$

\Rightarrow the projective transformations form a group. ■

12.2 Solutions to Exercises in Chapter 3

Exercises on Page 56

1. Given a line

$$\ell : x_2 = mx_1 + b,$$

we shall verify the corresponding point of ℓ ,

$$\bar{\ell} : \left(\frac{d}{1-m}, \frac{b}{1-m} \right), \quad m \neq 1.$$

Choose two arbitrary points on ℓ : $A_1 = (a_1, ma_1 + b)$ and $A_2 = (a_2, ma_2 + b)$.
The point A_1 is represented in \parallel -coords as the line

$$\bar{A}_1 : x_2 = \frac{ma_1 + b - a_1}{d}x_1 + a_1.$$

The point A_2 is represented in \parallel -coords as the line

$$\bar{A}_2 : x_2 = \frac{ma_2 + b - a_2}{d}x_1 + a_2.$$

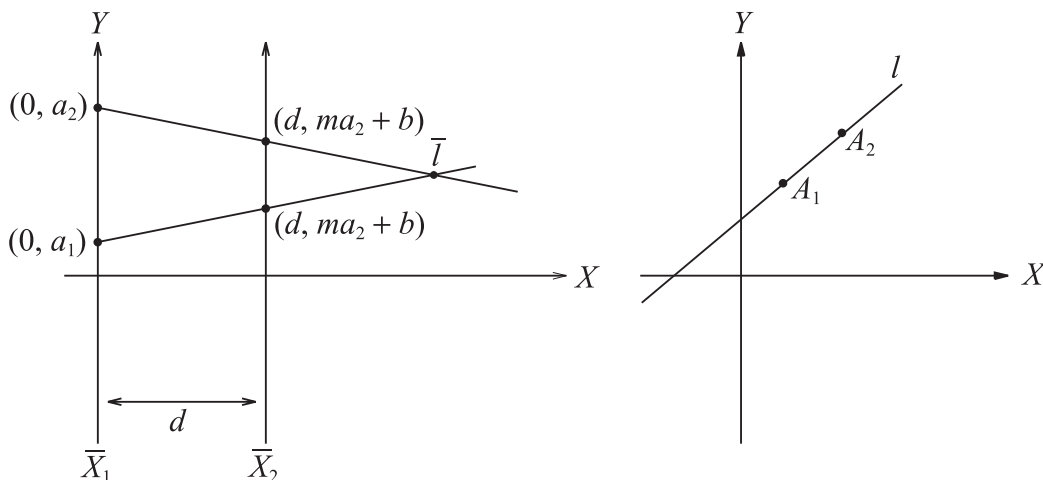
The intersection point of these lines is

$$\bar{\ell} : \left(\frac{d}{1-m}, \frac{b}{1-m} \right), \quad m \neq 1.$$

This point is not dependent on a_1 and a_2 . Therefore, every two points on ℓ are represented as two lines in the dual plane that always intersect at $\bar{\ell}$.

2. Let us check whether $[B^{-1}A]^k = I$:

$$B^{-1}A = \begin{bmatrix} 0 & -d & -1 \\ -d & -d & 0 \\ 0 & d^2 & 0 \end{bmatrix} \quad \forall k \in \mathbb{N} : [B^{-1}A]^k \neq I.$$



3. Let us check whether $[A^{-1}B]^k = I$:

$$A^{-1}B = \begin{bmatrix} 0 & -\frac{1}{d} & -\frac{1}{d^2} \\ 0 & 0 & \frac{1}{d^2} \\ -1 & 0 & -\frac{1}{d} \end{bmatrix} \quad \forall k \in \mathbb{N} : [A^{-1}B]^k \neq I.$$

There are some dualities that do not repeat themselves, and this is one of them.

Exercises on Page 61

1. Let us look at the point $P(-1, 1)$ and the line $\ell : y = 2x + 4$. The translation of the origin O to the point $(2, 5)$ is represented in \parallel -coords as a rotation. We can see that the coordinates are moved and rotated to become the dotted ones. The point $P(-1, 1) \rightarrow P'(-3, -4)$.

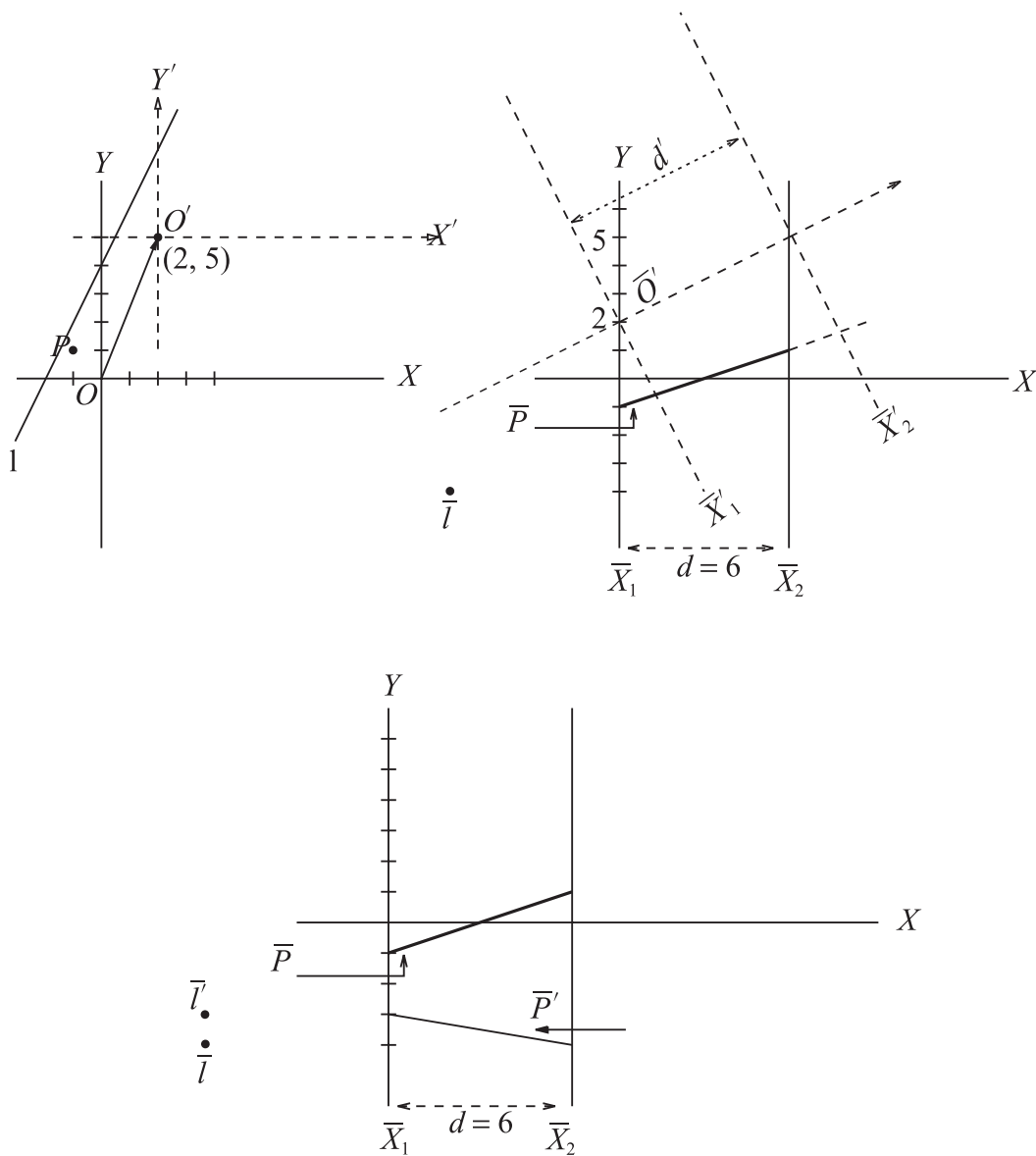
The line $\ell : y = 2x + 4 \rightarrow \ell' y = 2x + 3$.

In \parallel -coords $\bar{\ell} = (-6, -4) \rightarrow \bar{\ell}' = (-6, -3)$.

The distance d becomes d' :

$$d' = \sqrt{(y_1 - x_1)^2 + d^2}.$$

2. Without moving the coordinates, moving only the objects P and ℓ , the reflection about the x_1 and x_2 axes: ℓ is reflected to $\ell_1 : y = -2x - 4$ about the x axis and to $\ell_2 : y = -2x + 4$ about the y axis. In \parallel -coords, $\bar{\ell}_1 = (2, -\frac{4}{3})$ and $\bar{\ell}_2 = (2, \frac{4}{3})$. The point P is reflected to $P_1(-1, -1)$ about the x axis, and to $P_2(1, 1)$ about the y axis. For \parallel -coords this is shown in Figure 12.21.

**Figure 12.20.** Translation \leftrightarrow rotation.

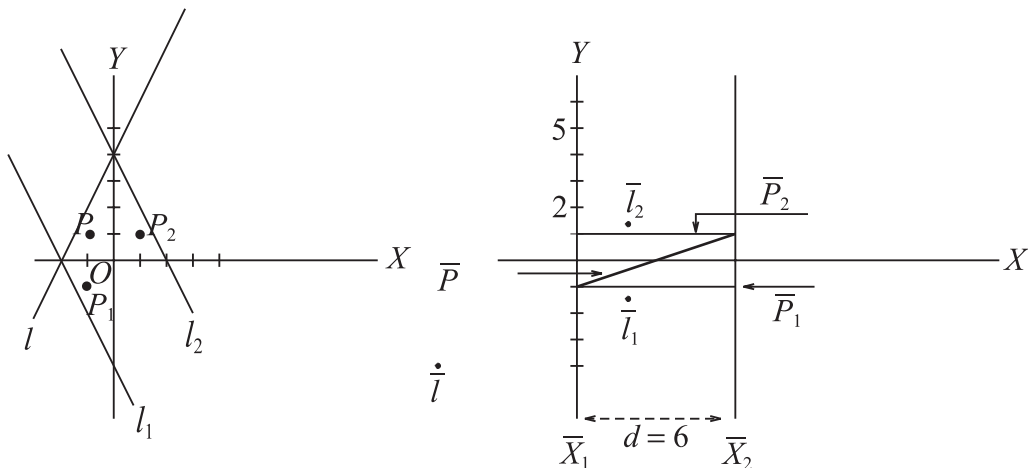


Figure 12.21. Reflection about x axis.

3. The rotation of the axes about the origin by an angle: θ ($m = \tan(\theta)$)
In \parallel -coords,

$$X'_1 = \left(\frac{d}{1 - \tan(90 + \theta)}, 0 \right), \quad X'_2 = \left(\frac{d}{1 - \tan(\theta)}, 0 \right).$$

Cartesian coordinates	\parallel -coords
$X\text{-axis} \longrightarrow X' : y = \tan \theta x$	$\bar{X}_1 \longrightarrow \bar{X}'_1 = \left(\frac{d}{1 - \tan \theta}, 0 \right)$
$Y\text{-axis} \longrightarrow Y' : y = \frac{-1}{\tan \theta} x$	$\bar{X}_2 \longrightarrow \bar{X}'_2 = \left(\frac{d}{1 + \frac{1}{\tan \theta}}, 0 \right)$

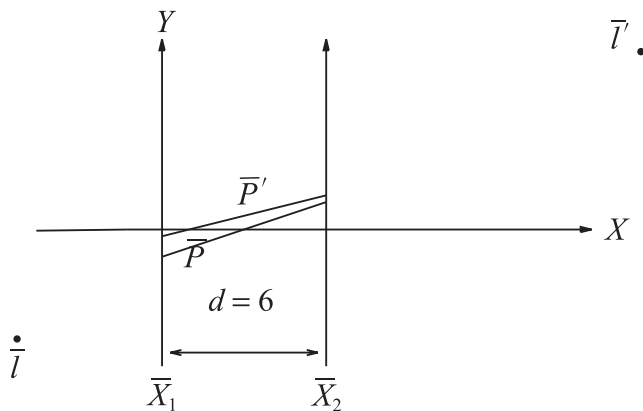
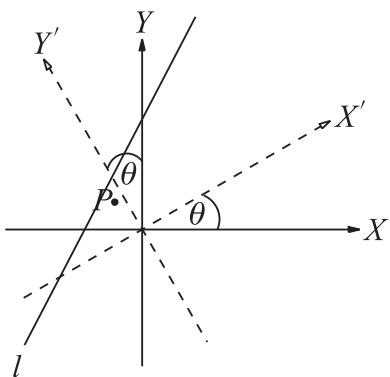
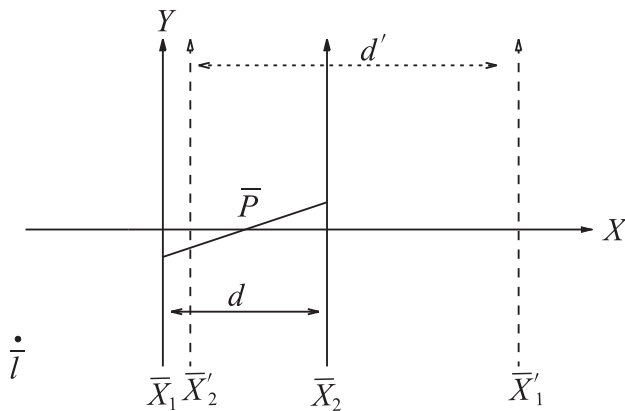
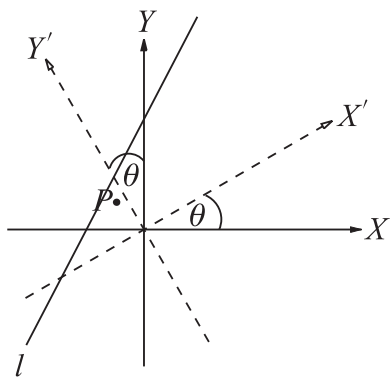
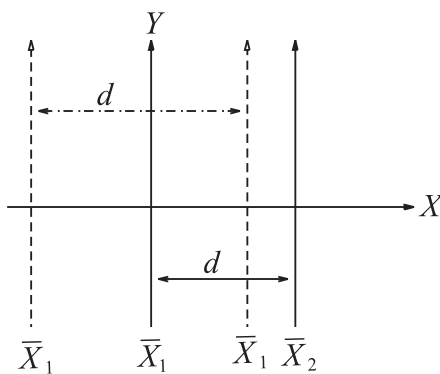
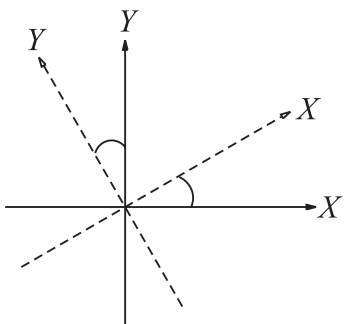
The distance d becomes d' :

$$\theta \neq 45, \quad d' = \left| \frac{d(1 + \tan^2 \theta)}{1 - \tan^2 \theta} \right|.$$

Choose $\theta = 30^\circ$. Then, $\bar{X}_1 \longrightarrow \bar{X}'_1(14.19, 0)$ and $\bar{X}_2 \longrightarrow \bar{X}'_2(2.19, 0)$. If we don't move the coordinates but only the objects P and ℓ , then

$$\begin{aligned} P' &= \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos 30 & \sin 30 \\ -\sin 30 & \cos 30 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.366 \\ 1.366 \end{pmatrix}, \end{aligned}$$

$$\ell' : y = 0.66x + 2.142 \longrightarrow \bar{\ell}' = (17.67, 6.309).$$



4. The composition of the translation of the origin to the point $(2, 5)$ followed by the rotation of the axes by angle 30° yields

$$P = (-1, 1) \longrightarrow P' = (-3, -4) \longrightarrow P'' = (-4.59, -1.96),$$

$$\ell : y = 2x + 4 \longrightarrow \ell' : y = 2x + 3 \longrightarrow \ell'' : y = 0.659x + 1.6,$$

$$\bar{\ell} = (-6, -4) \longrightarrow \bar{\ell}' = (-6, -3) \longrightarrow \bar{\ell}'' = (17.59, 4.69).$$

12.3 Solutions to Exercises in Chapter 4

Exercises on Page 73

1. Consider a line $\ell \subset \mathbb{R}^N$ whose description, by a set of linear equations, contains $x_i = c_i$, a constant, for some i . Such a line can be represented by $N - 1$ indexed points. Let S be the set of indices for which the variables are not constant, where $S = (i_1, \dots, i_M)$, and let \tilde{S} be the complementary set of indices, so

$$\ell_{ij} : x_j = m_i^j x_i + b_i^j \quad \forall i < j \in S = (i_1, \dots, i_M),$$

$$\ell_{0k} : x_k = b_k \quad \forall k \in \tilde{S} = (i_{M+1}, \dots, i_N).$$

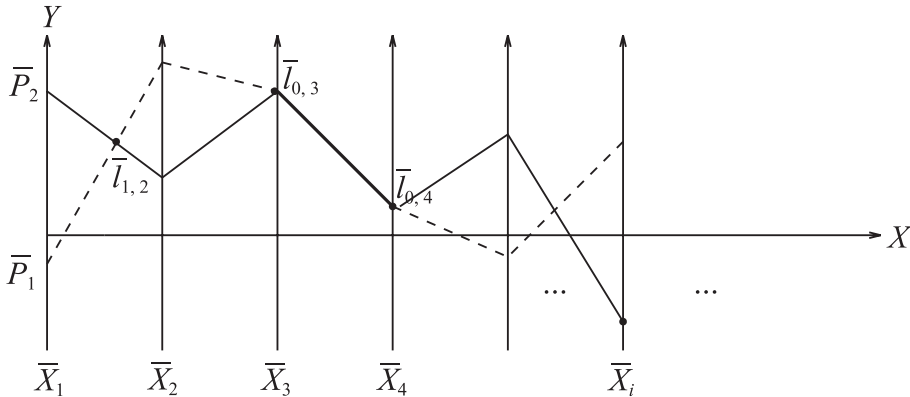
In homogenous coordinates,

$$\bar{\ell}_{ij} = (j - i, b_i^j, 1 - m_i^j) \quad \forall i, j \in S = (i_1, \dots, i_M); i \neq j,$$

$$\bar{\ell}_{0k} = (k - 1, b_k, 1) \quad \forall k \in \tilde{S} = (i_{M+1}, \dots, i_N).$$

2. Consider the plane

$$\pi_{123} : c_1 x_1 + c_2 x_2 + c_3 x_3 = c_0.$$



Let us prove that

$$\bar{\pi}_{123} = (c_2 + 2c_3, c_0, c_1 + c_2 + c_3).$$

The plane π_{123} intersects at least two of the planes $\pi_1 : x_1 = 0$, $\pi_2 : x_2 = 0$, and $\pi_3 : x_3 = 0$.

Let us find the representation of the intersection lines in homogeneous coordinates:

$$L_1 = \pi_{123} \cap \pi_1 \implies c_2x_2 + c_3x_3 = c_0 \implies \ell_{23} : x_3 = -\frac{c_2}{c_3}x_2 + \frac{c_0}{c_3},$$

$$\bar{\ell}_{23} = (2c_3 + c_2, c_0, c_3 + c_2) \text{ and } \bar{\ell}_{01} = (0, 0, 1) \text{ (const.)},$$

$$\begin{aligned} L_1 &= \left[\left| \begin{array}{cc|c} c_0 & c_3 + c_2 & \\ 0 & 1 & \end{array} \right|, \left| \begin{array}{cc|c} c_3 + c_2 & 2c_3 + c_2 & \\ 1 & 0 & \end{array} \right|, \left| \begin{array}{cc|c} 2c_3 + c_2 & c_0 & \\ 0 & 0 & \end{array} \right| \right] \\ &= [-c_0, 2c_3 + c_2, 0], \end{aligned}$$

$$L_2 = \pi_{123} \cap \pi_2 \implies c_1x_1 + c_3x_3 = c_0 \implies \ell_{13} : x_3 = -\frac{c_1}{c_3}x_1 + \frac{c_0}{c_3},$$

$$\bar{\ell}_{13} = (2c_3, c_0, c_3 + c_1) \text{ and } \bar{\ell}_{02} = (1, 0, 1) \text{ (const.)},$$

$$\begin{aligned} L_2 &= \left[\left| \begin{array}{cc|c} c_0 & c_3 + c_1 & \\ 0 & 1 & \end{array} \right|, \left| \begin{array}{cc|c} c_3 + c_1 & 2c_3 & \\ 1 & 1 & \end{array} \right|, \left| \begin{array}{cc|c} 2c_3 & c_0 & \\ 1 & 0 & \end{array} \right| \right] \\ &= [c_0, c_1 - c_3, -c_0], \end{aligned}$$

$$L_3 = \pi_{123} \cap \pi_3 \implies c_1x_1 + c_2x_2 = c_0 \implies \ell_{12} : x_2 = -\frac{c_1}{c_2}x_1 + \frac{c_0}{c_2},$$

$$\bar{\ell}_{12} = (c_2, c_0, c_2 + c_1) \text{ and } \bar{\ell}_{03} = (2, 0, 1) \text{ (const.)},$$

$$\begin{aligned} L_3 &= \left[\left| \begin{array}{cc|c} c_0 & c_2 + c_1 & \\ 0 & 1 & \end{array} \right|, \left| \begin{array}{cc|c} c_2 + c_1 & c_2 & \\ 1 & 2 & \end{array} \right|, \left| \begin{array}{cc|c} c_2 & c_0 & \\ 2 & 0 & \end{array} \right| \right], \\ &= [c_0, 2c_1 + c_2, -2c_0], \end{aligned}$$

$$L_1 \cap L_2 = \bar{\pi}_{123}$$

$$= \left(\left| \begin{array}{cc|c} 2c_3 + c_2 & 0 & \\ c_1 - c_3 & -c_0 & \end{array} \right|, \left| \begin{array}{cc|c} 0 & -c_0 & \\ -c_0 & c_0 & \end{array} \right|, \left| \begin{array}{cc|c} -c_0 & 2c_3 + c_2 & \\ c_0 & c_1 - c_3 & \end{array} \right| \right)$$

$$= (-c_0(2c_3 + c_2), -c_0^2, -c_0(c_1 + c_2 + c_3))$$

$$= (2c_3 + c_2, c_0, c_1 + c_2 + c_3),$$

$$L_1 \cap L_3 = \bar{\pi}_{123}$$

$$= \left(\left(\begin{array}{cc|c} 2c_3 + c_2 & 0 & \\ 2c_1 + c_2 & -2c_0 & \end{array} \right), \left(\begin{array}{cc|c} 0 & -c_0 & \\ -2c_0 & c_0 & \end{array} \right), \left(\begin{array}{cc|c} -c_0 & 2c_3 + c_2 & \\ c_0 & 2c_1 + c_2 & \end{array} \right) \right)$$

$$= (-2c_0(2c_3 + c_2), -2c_0^2, -2c_0(c_1 + c_2 + c_3))$$

$$= (2c_3 + c_2, c_0, c_1 + c_2 + c_3),$$

$$L_2 \cap L_3 = \bar{\pi}_{123}$$

$$= \left(\left(\begin{array}{cc|c} c_1 - c_3 & -c_0 & \\ 2c_1 + c_2 & -2c_0 & \end{array} \right), \left(\begin{array}{cc|c} -c_0 & c_0 & \\ -2c_0 & c_0 & \end{array} \right), \left(\begin{array}{cc|c} c_0 & c_1 - c_3 & \\ c_0 & 2c_1 + c_2 & \end{array} \right) \right)$$

$$= (c_0(2c_3 + c_2), c_0^2, c_0(c_1 + c_2 + c_3))$$

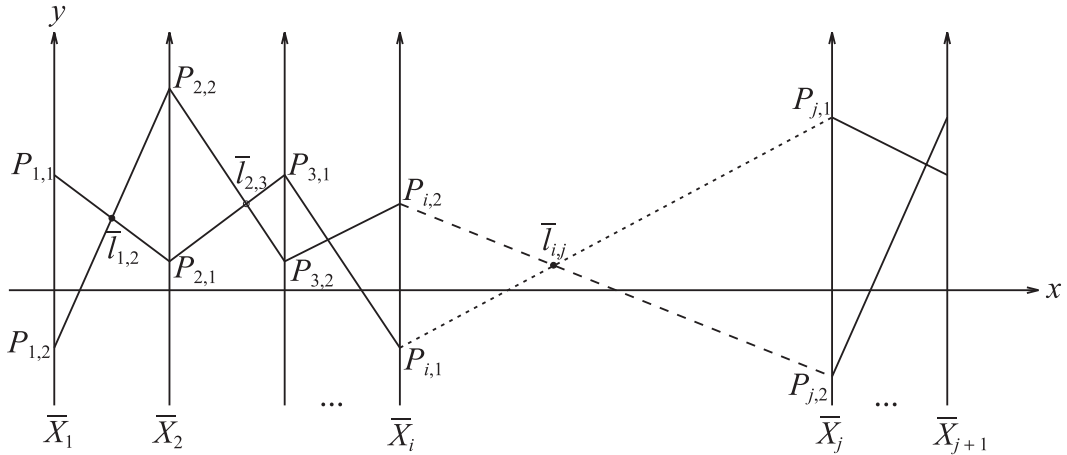
$$= (2c_3 + c_2, c_0, c_1 + c_2 + c_3) \quad \blacksquare$$

3. Algorithm for determining the point $\bar{\ell}_{ij}$, for any distinct pair i, j of a line ℓ given *all* the points $\bar{\ell}_{i-1,i}, i = 1, 2, \dots, N$:
- (a) Choose a point $P_{1,1}$ on the \bar{X}_1 axis.
 - (b) Construct the line through $P_{k,1}$ and $\bar{\ell}_{k,k+1}$ for $k = 1$. Define the intersection point with the \bar{X}_{k+1} axis as $P_{k+1,1}$. Repeat this step for $k = 2, \dots, N$. The polygonal line through the set of points $P_{k,1}$ defines a point on the original line ℓ .
 - (c) Choose another point $P_{1,2}$ on the \bar{X}_1 axis. Repeat step (b) for all the points $P_{k,2}, k = 1, \dots, N$. We get another polygonal line through $P_{k,2}, k = 1, \dots, N$.
 - (d) Construct the line L_1 through $P_{i,1}, P_{j,1}$ and the line L_2 through $P_{i,2}, P_{j,2}$. The intersection between L_1 and L_2 is the required point $\bar{\ell}_{i,j}$.

Complexity: $O(N)$.

4. For lines where some of the m 's are close to 1 we make a transformation of variables $x'_i = e_i x_i$, where $e_i = 1 \forall i \in \tilde{S}$ (for the constants). For the variables in S the linear equation is

$$\ell_{ij} : x_i = m_i^j x_j + b_j.$$



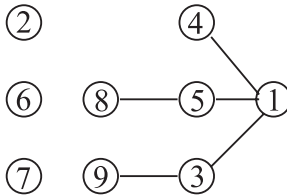
The remaining e_i are developed recursively. Starting at the root index i , set $e_i = 1$ and visit the remaining nodes in order, setting $e_j = +1$ to ensure that the slope m'_{ij} of the line

$$\ell'_{ij} : x_i = m'_{ij} x_j + b'_j$$

is negative. Here $m'_{ij} = e_i e_j m^j_i$ and $b'_j = e_i b_j$. Since the slope is negative, it follows that $\bar{\ell}'_{ij}$ is between \bar{X}_{i-1} and \bar{X}_{j-1} .

Exercises on Page 86

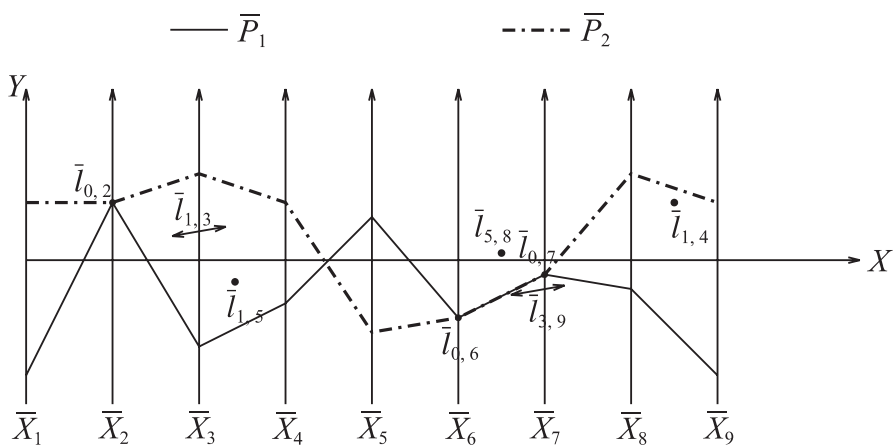
1. (a) For the example, choose as a start variable x_1 .



(b) The following table represents the lines ℓ_{ij} :

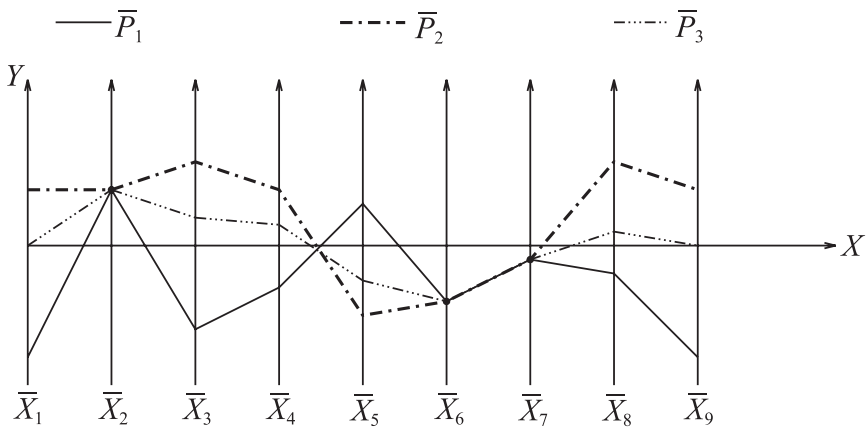
Coordinates	Point 1	Point 2	i	j	$x_{i,j}$	$y_{i,j}$	$h_{i,j}$	Point 3
X_1	-4	2	1	3	2	1	0	0
X_2	2	2	1	4	7.5	2	1	2
X_3	-3	3	1	5	2.4	-0.7	1	1
X_4	-1.6	2	3	9	6	-1	0	0.8
X_5	1.5	-2.5	5	8	5.5	0.25	1	$-\frac{7}{6}$
X_6	-2	-2	0	2	1	2	1	-2
X_7	-0.5	-0.5	0	6	5	-2	1	-0.5
X_8	-1	3	0	7	6	-0.5	1	$\frac{2}{3}$
X_9	-4	2						0

X_{13} and X_{39} are ideal points.

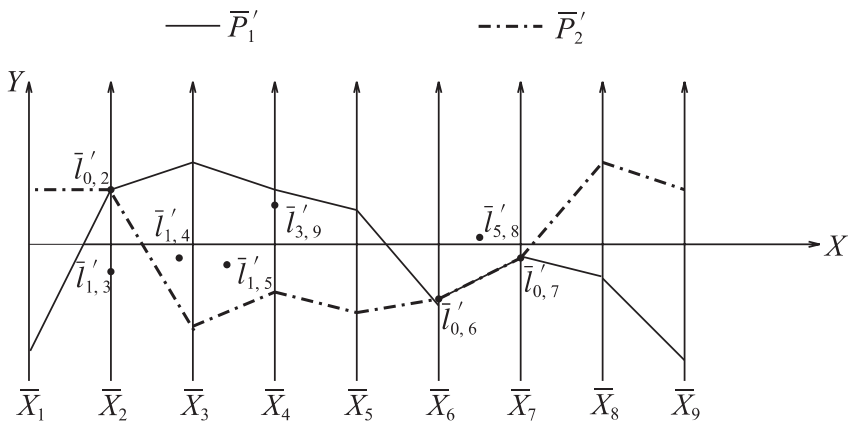


(c) Find a point P_3 on the same line segment between P_1 and P_2 .

(d)+(e) Find $\bar{\ell}'$ and the points P'_1 and P'_2 in the x' system:



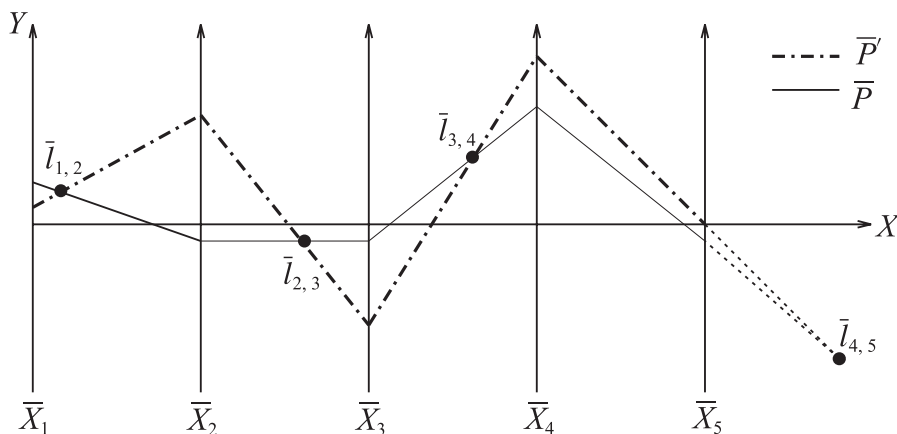
Coordinates	Point 1	Point 2	i	j	$x_{i,j}$	$y_{i,j}$
x'_1	-4	2	1	3	1	-0.5
x'_2	2	2	1	4	1.87	-0.25
x'_3	3	-3	1	5	2.4	-0.7
x'_4	2	-1.6	3	9	3	1.5
x'_5	1.5	-2.5	5	8	5.5	0.25
x'_6	-2	-2	0	2	1	2
x'_7	-0.5	-0.5	0	6	5	-2
x'_8	-1	3	0	7	6	-0.5
x'_9	-4	2				



2. This is an example of the tree enumeration for $N = 4$.
 - (a) Only four distinct kinds of trees are possible.
 - (b) The number of distinct rooted trees is shown below each tree. There are 64 distinct rooted trees: $N^{N-1} \rightarrow 4^{4-1} = 64$.

Exercise on Page 100

1. Let us show the translation of the point P to a new position P' and the line unrotated on P' :



12.4 Solutions to Exercises in Chapter 5

Exercises on Page 122

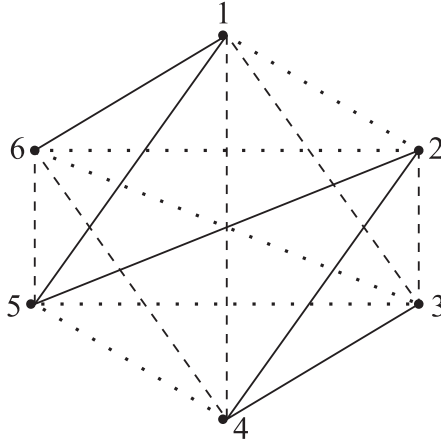
1. (a) For example, look at the undirected graph of $N = 6$. We have three permutations whose every two vertices are adjacent:

$$\begin{cases} P_1 \rightarrow (1, 2, 6, 3, 5, 4) \\ P_2 \rightarrow (2, 3, 1, 4, 6, 5) \\ P_3 \rightarrow (3, 4, 2, 5, 1, 6) \end{cases}$$

Notice that the relationship between the permutations is that each time, we added 1 (mod N) to the previous permutation. Notice that $6 + 1 \equiv 1(\text{mod } 6)$.

- (b) An algorithm for computing the $O\left(\frac{N}{2}\right)$ permutations:

Input: a two-way queue of N vertices (by order).



Output: if N is even, $\left(\frac{N}{2}\right) \times (N)$ matrix (called “Permut $[i] [j]$ ”), if N is odd, $\left(\frac{N}{2} + 1\right) \times (N)$ matrix; every line in the matrix is a permutation.

```

k = 1
Permut[1][k]=dequeue-head
Permut[1][k++]=dequeue-head
  while the queue is not empty do
    Permut[1][k++]=dequeue-head
    Permut[1][k++]=dequeue-tail
if N is even
  for i=2 to  $\frac{N}{2}$ 
    for j=1 to N
      Permut[i][j]=Permut[i - 1][j]+1 (mod N)
if N is odd
  for i=2 to  $\frac{N}{2} + 1$ 
    for j=1 to N
      Permut[i][j]=Permut[i - 1][j]+1 (mod N)

```

- (c) Research problem.
- (d) Research problem.
- 2. Given three noncollinear points in \mathbb{R}^3 , an algorithm for constructing a set of planar coordinates for the plane determined by the three points A, B, C is as follows:
 - i. First check whether one of the lines AB, BC, AC lies on one of the principal 2-planes; if so, this line determines a planar axis y^1 . In order to determine

the other planar axis, assume that AB lies on x_2x_3 and determine y^1 , and AC intersects plane x_1x_2 at a point D ; AB intersects x_1 at E , so the other planar axis y^2 is determined by DE .

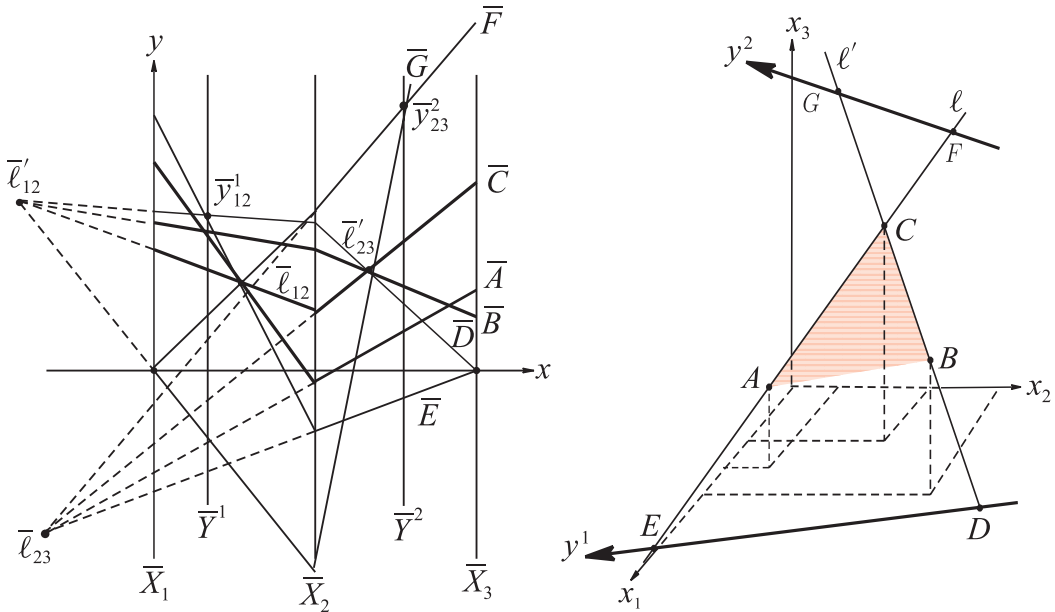
- ii. If two lines already lie on the principal 2-planes, they define both planar axes.
 - iii. If the three points have one common coordinate, for example, they have the same x_3 , it means that the plane is parallel to the x_1x_2 plane; find the point of intersection of the lines AB , AC , BC with the planes x_1x_3 and x_2x_3 to obtain four points: G and F on x_1x_3 determine y^1 , and D and M on x_2x_3 determine y^2 .
 - iv. Otherwise, if the points A , C determine the line ℓ and B , C determine the line ℓ' , the points are not collinear; therefore $\bar{\ell}_{IJ}$ and $\bar{\ell}'_{IJ}$ are not coincident. Find the intersecting points of ℓ and ℓ' with the x_1x_2 plane (points E and D) and with the x_2x_3 plane (points F and G). The lines DE and FG determine the y^1 and y^2 axes respectively. In parallel coordinates we get the points \bar{Y}_{12}^1 and \bar{Y}_{23}^2 , which complete the representation of the plane determined by A , B , and C .
3. Given a plane π parallel to one of the orthogonal axes, for example to the axis x_3 , an algorithm for constructing planar coordinates for the plane π , assuming $\pi \parallel x_3$, is as follows:
- i. If π intersects x_1 and x_2 , find the intersecting points with x_1 and x_2 (points A and B); they determine planar axes y^1 and $\bar{Y}^2 = \bar{X}_3$.
 - ii. If $\pi \parallel x_1$ as well, then find only the intersection point with x_2 ; this point will be the origin of the coordinate system, in \parallel -coords the planar axes coincident the ideal points \bar{X}_1 and \bar{X}_3 .
4. Given a set of planar coordinates y^1 and y^2 for the plane π , here is an algorithm for constructing a plane $\pi' \perp \pi$: rotate y^2 about y^1 by 90° . Next let

$$y^2 : \begin{cases} x_3 = m_3x_2 + b_3, \\ x_1 = 0. \end{cases}$$

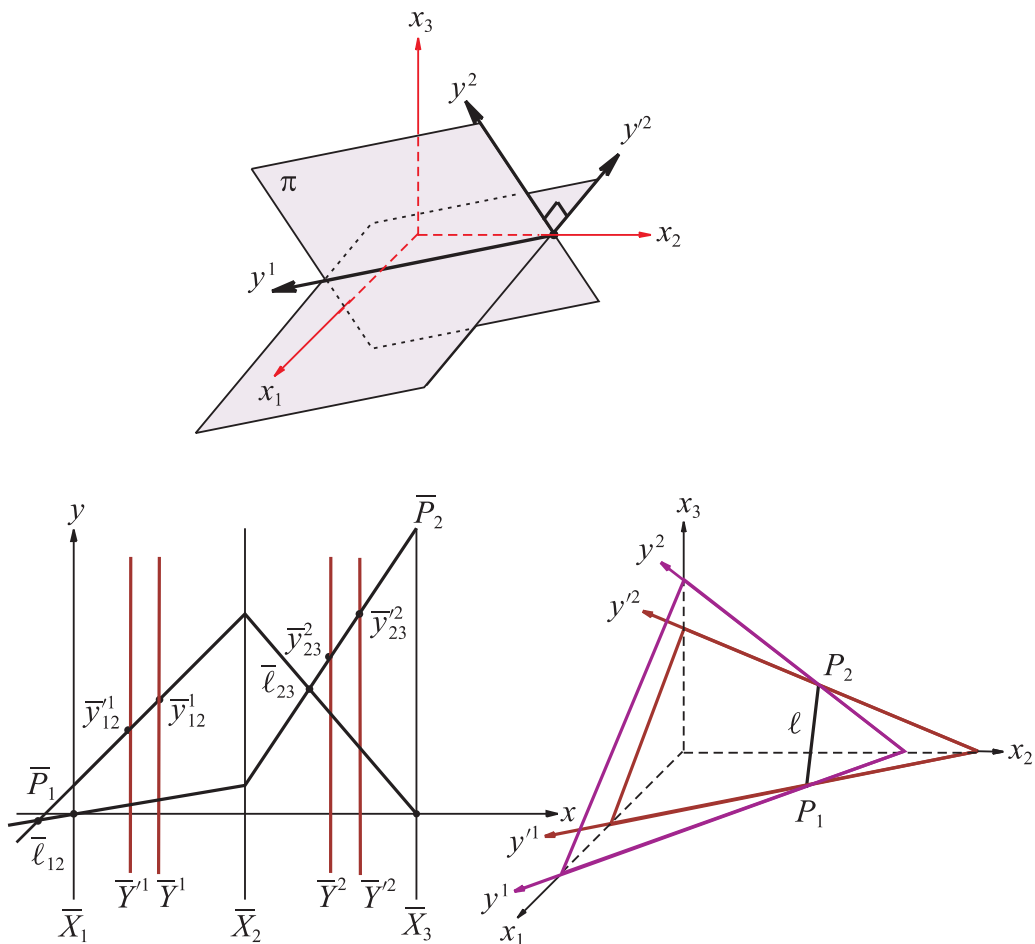
After rotating by 90° ,

$$y'^2 : \begin{cases} x_3 = -\frac{1}{m_3}x_2 - \frac{b_3}{m_3}, \\ x_1 = 0. \end{cases}$$

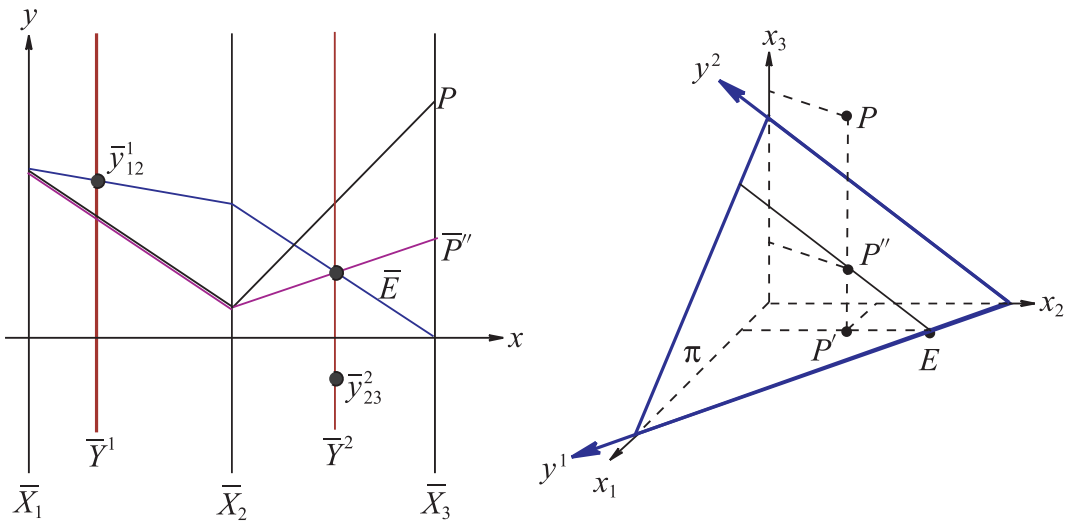
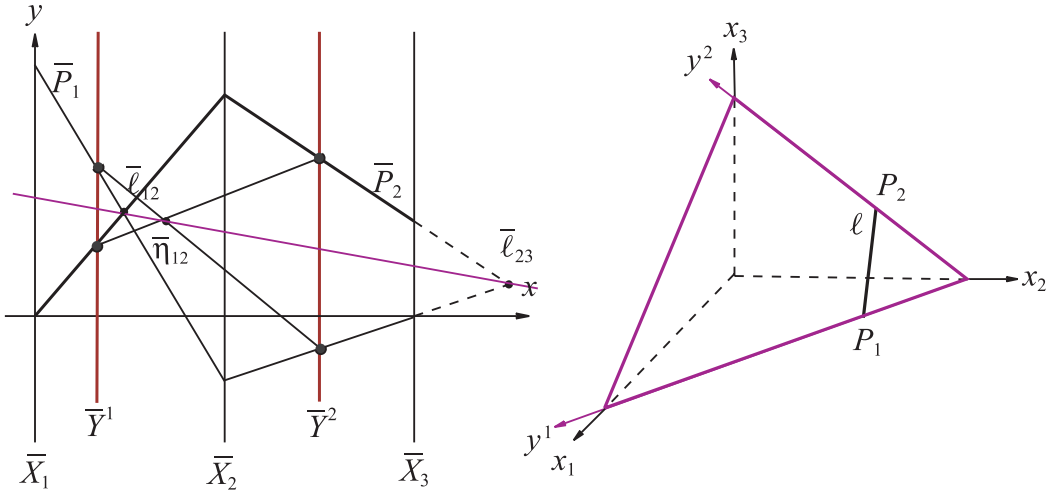
Then π' is determined by y^1 and y'^2 .



5. Here is an algorithm for determining the intersection line of two planes: given two planes π and π' with their planar coordinates y^1, y^2 and y'^1, y'^2 , find the intersecting points $P_1 = y^1 \cap y'^1$ and $P_2 = y^2 \cap y'^2$; the intersection line ℓ is determined by P_1 and P_2 . For a different construction, see Section 5.3.4.
6. Given a plane π and a line ℓ , here is an algorithm for determining whether $\ell \subset \pi$: in \parallel -coords π is represented by two axes \bar{Y}^1 and \bar{Y}^2 ; ℓ is represented by two points $\bar{\ell}_{12}$ and $\bar{\ell}_{23}$. If $\ell \subset \pi$, then ℓ can be represented, relative to the planar coordinates $y^1 y^2$, by a single point η_{12} . Let $P_1 = \ell \cap x_1 x_2$, $P_2 = \ell \cap x_2 x_3$; in \parallel -coords connect the intersections of P_1 with \bar{Y}^1 and \bar{Y}^2 with a straight line; the same is done with the intersections of P_2 with \bar{Y}^1 and \bar{Y}^2 . The intersection of these two straight lines is the point η_{12} . If the three points $\bar{\ell}_{12}, \bar{\ell}_{23}, \eta_{12}$ are collinear, then $\ell \subset \pi$.
7. Given a plane π and a point P , here is an algorithm for determining whether P is below or above π :
 - i. If π is not parallel to any of the orthogonal axes, let P' be the projection of P on the $x_1 x_2$ plane. Let P'' be the projection of P on the plane π . In \parallel -coords construct the polygonal line \bar{E} through \bar{y}_{23}^1 having the same x_1 coordinate as \bar{P} and $x_3 = 0$. Next construct the polygonal line \bar{P}'' having the same x_1 and x_2 as \bar{P} and the same y^2 as \bar{E} . If the x_3 coordinate of \bar{P} is higher than the x_3 coordinate of \bar{P}'' , then P is above π .



- ii. If π is parallel to one of the orthogonal axes, assume $\pi \parallel x_3$; let P' be the projection of P on x_1x_2 plane and let P'' be the point on y^1 with the same x_1 coordinate as P' . If the x_2 coordinate of \bar{P} is higher than the x_2 coordinate of \bar{P}'' , then P is above π .
 - iii. If π is parallel to one of the principal 2-planes, assume $\pi \parallel x_1x_3$; let A be the origin of y^1y^2 . If the x_2 coordinate of \bar{P} is higher than the x_2 coordinate of \bar{A} , then P is above π .
8. An algorithm for intersecting a line with a plane: it is convenient to describe ℓ by its projecting planes. Refer to the x_1x_2 and x_2x_3 projective planes by π_1 and π_2 respectively, and call their intersections with π PA and PB respectively.



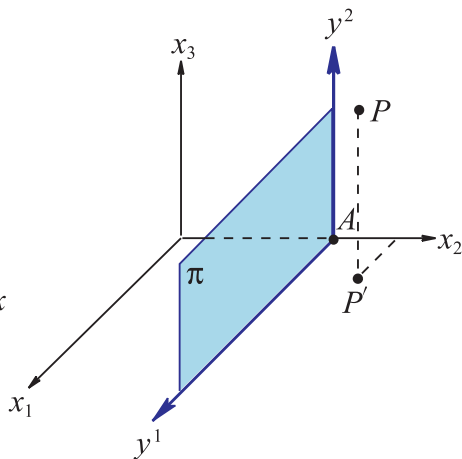
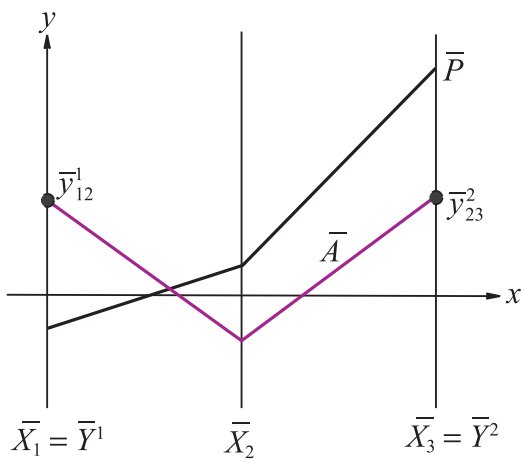
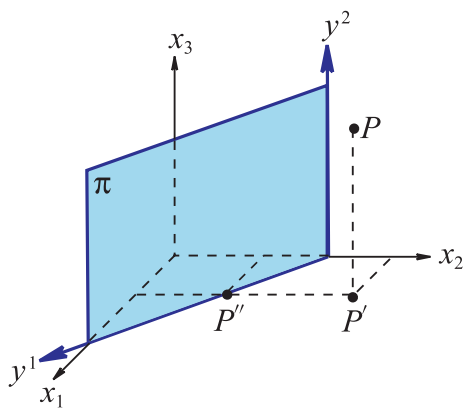
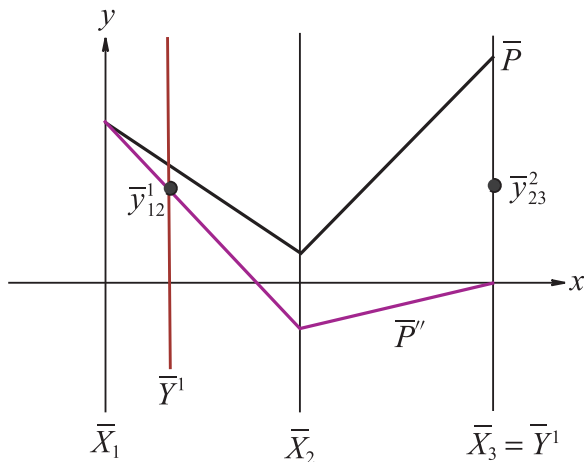
We see that

$$P = PA \cap \ell = PB \cap \ell = PA \cap PB \cap \ell.$$

It is easy to see that

$$\bar{\ell}_{12} = \overline{PA}_{12}, \quad \bar{\ell}_{23} = \overline{PB}_{23}$$

define the planar coordinates of π_1 and π_2 as z^1, z^2 and w^1, w^2 respectively. We can see that w^1 is coincident with the x_1 axis, z^2 is coincident with the



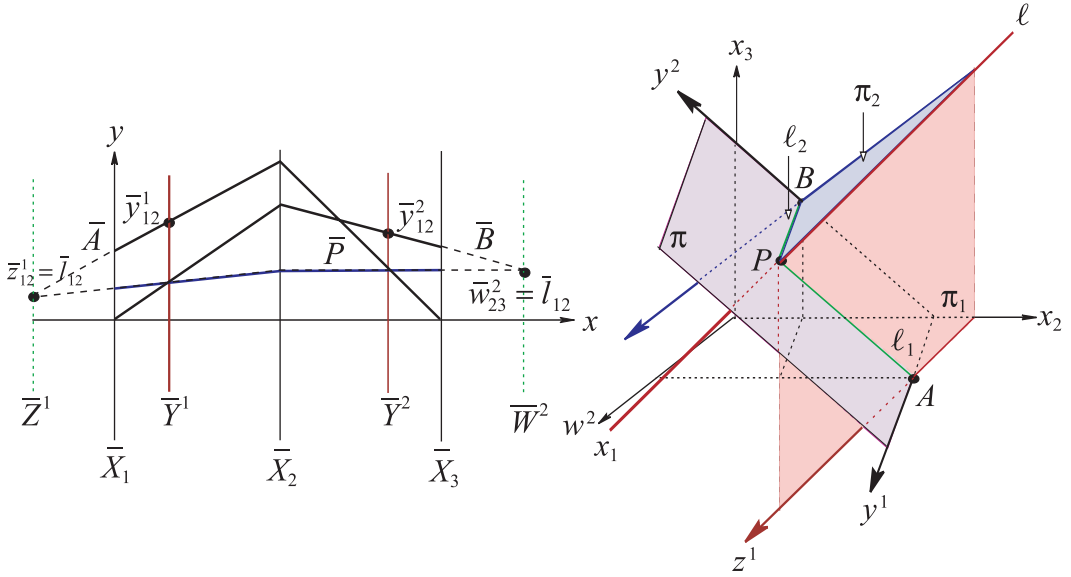
x_3 axis, $\bar{z}_{12}^1 = \bar{\ell}_{12}$, and $\bar{w}_{23}^2 = \bar{\ell}_{23}$. We next find the point $A = y^1 \cap z^1$ and $B = y^2 \cap w^2$; the polygonal line \bar{P} intersects \bar{A} on \bar{Y}^2 (because $AP \parallel y^2$) and intersects \bar{B} on \bar{Y}^1 (because $PB \parallel y^1$).

9. (a) i. A pencil of planes containing

$$\ell : \begin{cases} \ell_{12} : x_2 = m_2x_1 + b_2, \\ \ell_{23} : x_3 = m_3x_2 + b_3 \end{cases}$$

is represented by

$$\pi : (x_3 - m_3x_2 - b_3) + k(x_2 - m_2x_1 - b_2) = 0.$$



Proof. Every plane containing ℓ contains the points $(0, b_2, m_3b_2 + b_3)$ and $(-\frac{b_2}{m_2}, 0, b_3)$ and another general point (e, f, g) . We find that π is represented by

$$m_2 \left(\frac{fm_3 + b_3 - g}{b_2 - f + em_2} \right) x_1 - \left(\frac{m_3b_2 - g + b_3 + em_2m_3}{b_2 - f + em_2} \right) x_2 + x_3 + \left(\frac{m_3b_2 - g + b_3 + em_2m_3}{b_2 - f + em_2} \right) b_2 - m_3b_2 - b_3 = 0.$$

Let

$$K = \frac{g - fm_3 - b_3}{b_2 - f + em_2}.$$

Then

$$-m_2Kx_1 + (K - m_3)x_2 + x_3 - b_3 - Kb_2 = 0,$$

which is the required relation.

ii. Every plane represented by π contains ℓ .

Proof. Clearly, when $x_2 = m_2x_1 + b_2$, then $x_3 = m_3x_2 + b_3$ and conversely. So $\ell \subset \pi$.

- (b) Given $\pi : (x_3 - m_3x_2 - b_3) + k(x_2 - m_2x_1 - b_2) = 0$, therefore every line $\ell \subset \pi$ is represented by two linear equations:

$$\ell : \begin{cases} \ell_{12} : x_2 = m_2x_1 + b_2, \\ \ell_{23} : x_3 = m_3x_2 + b_3. \end{cases}$$

Let $y^1 = \pi \cap \{x_3 = 0\}$. Thus

$$\begin{aligned} y^1 : \begin{cases} y_{12}^1 : x_2 = \frac{km_2}{k-m_3}x_1 + \frac{kb_2+b_3}{k-m_3}, \\ y_{03}^1 : x_3 = 0 \end{cases} \\ \Rightarrow \bar{Y}^1 : \begin{cases} \bar{y}_{12}^1 = (k - m_3, kb_2 + b_3, k - m_3 - km_2), \\ \bar{y}_{03}^1 = (2, 0, 1). \end{cases} \end{aligned}$$

Let $y^2 = \pi \cap \{x_1 = 0\}$. Thus

$$\begin{aligned} y^2 : \begin{cases} y_{23}^2 : x_3 = (m_3 - k)x_2 + b_3 + kb_2, \\ y_{01}^2 : x_1 = 0 \end{cases} \\ \Rightarrow \bar{Y}^2 : \begin{cases} \bar{y}_{23}^2 = (2 + k - m_3, b_3 + kb_2, 1 + k - m_3), \\ \bar{y}_{01}^2 = (0, 0, 1). \end{cases} \end{aligned}$$

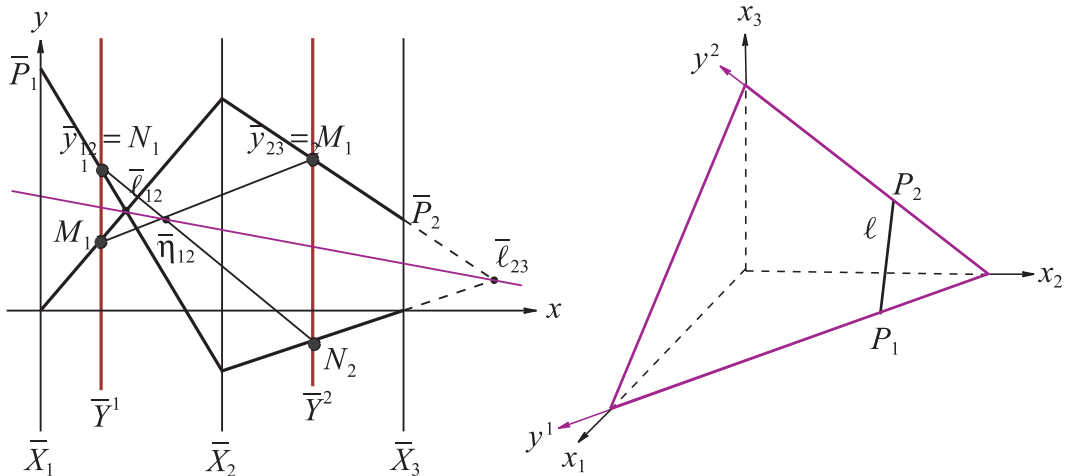
Let

$$N_1 = \bar{P}_1 \cap \bar{Y}^1 = \bar{y}_{12}^1 = (k - m_3, kb_2 + b_3, k - m_3 - km_2),$$

$$N_2 = \bar{P}_1 \cap \bar{Y}^2 = (m_3(2 + k - m_3), b_3m_3 - b_3k, m_3(1 + k - m_3)).$$

Construct the line through N_1 and N_2 :

$$\begin{aligned} \text{line1} &= \left[\begin{vmatrix} (N_1)_2 & (N_1)_3 \\ (N_2)_2 & (N_2)_3 \end{vmatrix}, \begin{vmatrix} (N_1)_3 & (N_1)_1 \\ (N_2)_3 & (N_2)_1 \end{vmatrix}, \begin{vmatrix} (N_1)_1 & (N_1)_2 \\ (N_2)_1 & (N_2)_2 \end{vmatrix} \right] \\ &= [kb_2m_3 + m_3k^2b_2 - kb_2m_3^2 + b_3m_3 - m_3b_3k + b_3k^2 \\ &\quad + km_2b_3m_3 - k^2m_2b_3, \\ &\quad m_3(-m_3 + km_2m_3 + k - 2km_2 - k^2m_2), \\ &\quad m_3b_3k - b_3k^2 - 2kb_2m_3 - m_3k^2b_2 + kb_2m_3^2 - 2b_3m_3]. \end{aligned}$$

$$\begin{aligned} M_1 &= \bar{P}_2 \cap \bar{Y}^1 = (k - m_3, b_2k - b_2m_3, k - m_3 - km_2), \\ M_2 &= \bar{P}_2 \cap \bar{Y}^2 = \bar{y}_{23}^2 = (2 + k - m_3, b_3 + kb_2, 1 + k - m_3). \end{aligned}$$
$$\begin{aligned} \text{line2} &= \left[\left| \begin{pmatrix} (M_1)_2 & (M_1)_3 \\ (M_2)_2 & (M_2)_3 \end{pmatrix}, \left| \begin{pmatrix} (M_1)_3 & (M_1)_1 \\ (M_2)_3 & (M_2)_1 \end{pmatrix}, \left| \begin{pmatrix} (M_1)_1 & (M_1)_2 \\ (M_2)_1 & (M_2)_2 \end{pmatrix} \right| \right] \\ &= [kb_2 - kb_2m_3 - b_2m_3 + b_2m_3^2 - b_3k + b_3m_3 \\ &\quad + k^2m_2b_2 + km_2b_3, \\ &\quad - m_3 + km_2m_3 + k - 2km_2 - k^2m_2, \\ &\quad - (m_3 - k)(b_2m_3 + b_3 - 2b_2)], \\ \bar{\eta} &= \text{line1} \cap \text{line2} \text{ (in homogeneous coordinates),} \\ \bar{\eta} &= [-2m_3 + m_3^2 - k^2, -b_3m_3 - k^2b_2, (m_2 - 1)k^2 + m_3^2 - m_3] \end{aligned}$$


This page intentionally left blank

Notation and List of Symbols

Alphabetic

$\aleph_0, \aleph_1, \aleph_2, \dots$	cardinality of infinite sets
ℓ, r, s, \dots	lines are usually denoted by lowercase italics Latin letters
A, B, C, \dots	points are usually denoted by uppercase Latin letters
π, ρ, σ, \dots	planes and surfaces are usually denoted by lowercase Greek letters
\mathbb{N}	positive integers; natural numbers
\mathbb{R}	real numbers
\mathbb{C}	complex numbers
\Re	real part of a complex number
\Im	imaginary part of a complex number
\mathbb{R}^2	Euclidean 2-dimensional plane (or space)
\mathbb{P}^2	projective 2-dimensional plane (or space)
\mathbb{R}^N	Euclidean N-dimensional space
\mathbb{P}^N	projective N-dimensional space

Symbols

General

\in	belongs to
\ni	such that
\exists	there exists
\forall	for all
\subset	subset of
\cup	set union
\cap	set intersection

\emptyset	empty st (null set)
2^A	power set of A , the set of all subsets of A
ip	inflection point of a curve
\mathbf{n}	principal normal vector to a space curve
$\hat{\mathbf{n}}$	unit normal vector to a surface
\mathbf{t}	vector tangent to a space curve
\mathbf{b}	binormal vector
τ	torsion (measures “non-planarity” of a space curve)
\times	cross product
\rightarrow	mapping symbol
$A \rightarrow B$	correspondence (also used for duality)
\Rightarrow	implies
\Leftrightarrow	if and only if
$<, >$	less than, greater than
$ a $	absolute value of a
$ \mathbf{b} $	length of vector \mathbf{b}
$\ \ $	norm, length
\mathcal{L}_1	L_1 norm
\mathcal{L}_2	L_2 norm, Euclidean distance
\sum	summation
J_p	set of integers modulo p with the operations $+$ and \times
\approx	approximately equal
\neq	not equal
\parallel	parallel
\perp	perpendicular, orthogonal
$(\ , \ , \)$	homogeneous coordinates — ordered triple representing a point in \mathbb{P}^2 ; point coordinates
$[\ , \ , \]$	homogeneous coordinates — ordered triple representing a line \mathbb{P}^2 ; line coordinates
$p(A, B, C, D, \dots) \overset{P}{\underset{\wedge}{=}}$	perspectivity between the lines p and p' with respect to a point P
$p'(A', B', C', D', \dots)$	

$p \bar{\wedge} p'$	projectivity between the lines p and p'
$\frac{df}{dx}$	derivative of f with respect to x , also written as $f'(x)$
$\partial\Omega$	boundary of a region Ω
$\frac{\partial F}{\partial x_i}$	partial derivative of F with respect to x_i , also denoted F_{x_i} and when context is clear, F_i
∇F	gradient of F

Specific to Parallel Coordinates (In the order occuring in the text)

\parallel -coords	abbreviation for parallel coordinate system, parallel coordinates
\bar{A}	image of A in \parallel -coords, representation of A
$\bar{X}_1, \dots, \bar{X}_N$	axes for parallel coordinate system for N -dimensional space (either Euclidean or Projective)
$\bar{\ell}_{i,j}$	point representing the linear relation between x_i and x_j ; $(N - 1)$ points with two indices appear in the representation of a line in \mathbb{R}^N (or \mathbb{P}^N)
π^p	p -flat in \mathbb{R}^N , a plane of dimension $1 \leq p \leq (N - 1)$ in \mathbb{R}^N
π^1	a one-flat is a line, also denoted by ℓ
\mathbf{sp}	abbreviation for <i>superplane</i> , a 2-flat whose points are represented by straight lines in \parallel -coords
π^s	denotes a <i>superplane</i>
π^{Ns}	denotes a <i>superplane</i> in \mathbb{R}^N
\mathcal{E}	family of <i>superplanes</i> , named after J. Eickemeyer, who first discovered them
\mathcal{E}	also denotes the class of surfaces who are the envelope of their tangent (hyper) planes
\mathcal{EG}	class of surfaces contained in \mathcal{E} whose representing contours can be obtained by Lemma 9.3.1
d_i	directed distance of \bar{X}_i from the y-axis
\mathbf{d}_N^0	(see next), standard axis spacing, when the \bar{X}_1 is coincident with the y-axis and inter-axis distance is one unit for all axes
\mathbf{d}_N^i	axes spacing where the \bar{X}_1 axis is placed one unit after the \bar{X}_N and all axes up to and including \bar{X}_i are placed successively one unit apart and the rest of axes remain in the standard axes spacing.

π_1^s	first <i>superplane</i> , with the <i>standard axes spacing</i>
$\pi_{1'}^s$	<i>second superplane</i> , the \bar{X}_1 placed one unit to the right of \bar{X}_N the remaining axes are in <i>standard axes spacing</i> .
$\pi_{i'}^s$	is the i th <i>superplane</i> with axes spacing \mathbf{d}_N^i
$\bar{\pi}_{123}, \bar{\pi}_{1'123}$	two triple indexed points representing the plane (2-flat) $\pi \subset \mathbb{R}^3$, also denoted by $\bar{\pi}_{123} = \bar{\pi}_{0'}$, $\bar{\pi}_{1'23} = \pi_{1'}$ or 123, 1'23 when π is clear from the context; $\bar{\pi}_{1'23}$ is based on the axes spacing \mathbf{d}_3^1
$\bar{\pi}_{123, \dots, N} = \bar{\pi}_{0'}$	point with $(N + 1)$ indices representing a hyperplane (of dimension $N - 1$), and similarly $\bar{\pi}_{i'}$ for the representing points based on axes spacing \mathbf{d}_N^i
<i>gconics</i>	generalized conics, sections of a double cone in \mathbb{R}^3 whose 2 bases are the same bounded convex set in \mathbb{R}^2
<i>bc, ubc, gh</i>	like conics there are 3 kinds of gconics: bounded convex set (bc), unbounded convex set (ubc) and generalized hyperbola (gh)
\sqcup	outer union
\sqcap	inner intersection

Bibliography

- [1] A.E. Abbot. *A Romance of Many Dimensions*. Dover, New York (originally published in 1884 as 'Flatland by A Square', Seeley & Co., London), 1992.
- [2] N. Adrienko and G. Adrienko. *Constructing Parallel Coordinates Plots for Problem Solving*. in Proc. 1st Inter. Symp. on Smart Graphics, A. Butz, A. Krueger, P. Oliver and M. Zhou (eds.), ACM Press 9–14, New York, 2001.
- [3] C. Ahlberg and B. Shneiderman. *Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays*. Proc. CH94 Conf., 313–317, ACM Press, New York, 1994.
- [4] D. Asimov. The grand tour: A tool for viewing multidimensional data. *SIAM J. Sci. & Stat. Comp.*, 6:128–143, 1985.
- [5] T. Avidan and S. Avidan. Parallax – a data mining tool based on parallel coordinates. *Comput. Statist.*, 13–1:65, 1998.
- [6] S. Axelson. *Understanding Intrusion Detection Through Visualization*. Ph.D. Thesis, CS Dept. Chalmers Univ., Gotesborg, Sweden, 2005.
- [7] F. Ayres Jr. *Projective Geometry*. Shaum's Outlines on Math., McGraw-Hill, New York, 1967.
- [8] T.F. Banchoff. *Beyond the Third Dimension*. Scientific American Library, New York, 1990.
- [9] E.W. Bassett. Ibm's ibm fix. *Industrial Computing*, 14(41):23–25, 1995.
- [10] A.F. Beardon. *The Geometry of Discrete Groups*. Springer-Verlag, New York, 1983.
- [11] B. Bederson and B. Shneiderman. (eds.) *The Craft of Information Visualization: Readings and Reflection*. Morgan-Kaufman, San Francisco, 2003.
- [12] J. Bergen, J. Borwein and P. Borwein. *Pi: A Source Book*. Springer-Verlag, New York, 1997.
- [13] M. Berger. *Geometry I & II*. Corrected second printing, Springer-Verlag, New York, 1994.
- [14] J. Bertin. *Sémiologie Graphique*. First pub. 1967, Ecole des Hautes Etudes en Science, Paris, 1999.

- [15] D.M. Bloom. *Linear Algebra and Geometry*. Cambridge University Press, Cambridge, 1979.
- [16] B. Bollobas. *Graph Theory*. Springer-Verlag, New York, 1979.
- [17] V.G. Boltyanskii. *Envelopes*, R.B. Brown, translator (original in Russian). Pergamon Press, New York, 1964.
- [18] M.F. Brannan, D.A. Esplen and J.J. Gray. *Geometry*. Cambridge University Press, New York, 1999.
- [19] O.S. Brodetsky. *A First Course in Nomography (first published in 1920)*. G. Bell and Sons, London, 1949.
- [20] S. Bulthoff and H.H. Edelman. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Nat. Acad. Sci.*, 89-Psychology: 60–4, 1992.
- [21] J. Canny and J. Reif. *New Lower Bound Techniques for Robot Motion Planning Problems*, in *Proc. of 28th Symp. on Found. of Comp. Sci.* IEEE Comp. Soc., Washington, D.C., 1987.
- [22] S. Card, J. MacKinlay and B. Shneiderman. (eds.) *Readings in Information Visualization*. Morgan Kaufman, San Francisco, 1999.
- [23] A. Chatterjee. *Visualizing Multidimensional Polytopes and Topologies for Tolerances*. Ph.D. thesis, Dept. Comp. Sci., Univ. of S. Calif., 1995.
- [24] A. Chatterjee, P.P. Das and S. Bhattacharya. Visualization in linear programming using parallel coordinates. *Pattern Recognition*, 26–11:1725–36, 1993.
- [25] C. Chen. *Information Visualization: Beyond the Horizon, 2nd ed.* Springer, New York, 2004.
- [26] C.H. Chen. Generalized association plots for information visualization: The application of the convergence of iteratively formed correlation matrices. *Statistica Sica*, 12:1–23, 2002.
- [27] C.H. Chen and J.A. Chen. Interactive diagnosis plots for multidimensional scaling with applications in psychosis disorder data analysis. *Statistica Sica*, 10:665–691, 2000.
- [28] E.H. Chi. *A Framework for Visualizing Information*. Kluwer, Amsterdam, 2002.
- [29] H. Choi and H. Lee. *PCAV: Internet Attack Visualization in Parallel Coordinates*, LNCS 3783, 454–466. Springer-Verlag, New York, 2005.
- [30] T. Chomut. *Exploratory Data Analysis in Parallel Coordinates*. M.Sc. thesis, Dept. Comp. Sci., UCLA, 1987.
- [31] S. M. Cohan and D.C.H. Yang. Mobility analysis in parallel coordinates. *J. Mech. & Mach.*, 21:63–71, 1986.

- [32] G. Conti. *Security Data Visualization*. No Starch Press, San Francisco, 2007.
- [33] D. Conversano, D. Di Benedetto and R. Siciliano. *The Clockworktree: Visual Multivariating Splitting*. Cladag 2003, Book of Short Papers 113–7, Clueb Publisher, Bologna, 2003.
- [34] D. Cook and A. Buja. Manual controls for high-dimensional data projections. *J. of Comp. & Graph. Stat.*, 6–4, 464–480, 1997.
- [35] R. Courant. *Differential and Integral Calculus vol. 2*. Blackie & Son, London 171–183, 1946.
- [36] R. Courant and H.E. Robins. *What is Mathematics?* Oxford University Press, New York, 1953.
- [37] D. Cox, J. Little and D. O’Shea. *Ideals, Varieties, and Algorithms*. 2nd edition, Springer-Verlag, New York, 1997.
- [38] H.S.M. Coxeter. *Introduction to Geometry*. John Wiley, New York, 1969.
- [39] H.S.M. Coxeter. *Projective Geometry*. Second Ed., John Wiley, New York, 1974.
- [40] D.H. Deterding. *Speaker Normalization for Automatic Speech Recognition*. Ph.D. thesis, Cambridge University, 1989.
- [41] B. Dimsdale. *Conic Transformations – IBM LASC Tech. Rep. G320-2713*. IBM LA Scientific Center, 1981.
- [42] B. Dimsdale. *Conic Transformations and Projectivities – IBM LASC Tech. Rep. G320-2753*. IBM LA Scientific Center, 1984.
- [43] J.A. Dykes, A.M. MacEachren and M.J. Kraak. (eds). *Exploring Geo-visualization*. Elsevier, Amsterdam, 2005.
- [44] J.A. Dykes and D.M. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Comput. Statit. & Data Anal.*, 43 (Data Visualization II special ed.):581–603, 2003.
- [45] R. Edsall. The parallel coordinate plot in action: Design and use for geographic visualization. *Comput. Statit. & Data Anal.*, 43–4:605–619, 2003.
- [46] P. Ehrenfest. *In What Way Does it Become Manifest in the Fundamental Laws of Physics that Space has Three Dimensions? – Appeared in 1917 – in Collected Scientific Papers of P. Ehrenfest, Klein, M.J. ed., 400–409*. North-Holland Publishing Company, Interscience Publishers, New York, 1959.
- [47] J. Eickemeyer. *Visualizing p -flat in N -space using Parallel Coordinates*. Ph.D. thesis, Dept. Comp. Sci., UCLA, 1992.

- [48] P.L. Eisenhart. *A Treatise on the Differential Geometry of Curves and Surfaces*. Dover Publications, New York, 2004.
- [49] S. El Mejdani, R. Egli and F. Dubeau. Old & new straight-line detectors, descriptions & comparisons. *Pattern Recognition*, 41:1845–66, 2008.
- [50] A. Ellis and G. Dix. Enabling automatic clutter reduction in parallel coordinates. *Trans. on Vis. & Comp. Grap.*, 12–5:717–724, 2006.
- [51] N. Elmquist, J. Stasko and P. Tsingas. Datameadow: A visual canvas for analysis of large-scale multivariate data. *Proc. IEEE Vast Conf.*, to appear, 2007.
- [52] L. Ernstrom. A plücker formula for the singular projective varieties. *Commut. Algebra.*, 5–9:2897–2901, 1997.
- [53] K. Esbensen, et al. *Multivariate Analysis*, 1994, Camo AS, Trondheim, Norway.
- [54] S. Even. *Graph Algorithms*. Computer Science Press, Rockville MD, 1979.
- [55] M.G. Ewing. On the definition of inflection point. *Amer. Math. Monthly*, 45: 681–683, 1938.
- [56] F. Fallside and A.J. Robinson. *A Dynamic Connectionist Model of Phoneme Recognition*. Proc. of 1st European Neural Network Conf. (nEURO), 1988.
- [57] U.M. Fayad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge Mass., 1996.
- [58] U.M. Fayad, G. Grinstein and A. Wierse. *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann, San Francisco, 2001.
- [59] R. Finsterwalder. *A Parallel Coordinate Editor as a Visual Decision Aid in Multi-Objective Concurrent Control Engineering Environment* 119–122. IFAC CAD Contr. Sys., Swansea, UK, 1991.
- [60] P. Fiorini and A. Inselberg. *Configuration Space Representation in Parallel Coordinates*. IEEE Conf. Rob. & Aut., 1215–1220, 1989.
- [61] G. Fischer. *Plane Algebraic Curves*. Amer. Math. Soc. Press, Providence, 2001.
- [62] M. Friendly. Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *J. of Comp. & Graph. Stat.*, 8–3:373–95, 1999.
- [63] M. Friendly and al. *Milestones in Thematic Cartography*. www.math.yorku.ca/scs/SCS/Gallery/milestones/, 2005.
- [64] LORAL FSD. *Test Report for the Quick-Look Flight Demonstration of the IntraFormation Positioning System, Document Nos. 93-A37-002, Contract f33615-90-C-3609*. LORAL Federal Systems Co, Oswego, 1994.

- [65] M. O. Fua, Y.H. Ward and E.A. Rundensteiner. *Hierarchical Parallel Coordinates for Exploration of Large Datasets – Proc. of Conf. on Vis.*, 43–50. IEEE Comp. Soc. Press, Los Alamitos, CA, 1999.
- [66] G.W. Furnas and A. Buja. Prosection views: Dimensional inference through sections and projections. *J. Comp. & Graph. Stat.*, 3:323–85, 1994.
- [67] C. Gennings, K.S. Dawson, W.H. Carter and R.H. Myers. Interpreting plots of a multidimensional dose-response surface in parallel coordinates. *Biometrics*, 46:719–35, 1990.
- [68] W.J. Gibbs. Graphical methods in the thermodynamics of fluids. *Conn. Acad.*, II-1:309–342, 1873.
- [69] W.J. Gibbs. A method of geometrical representation of the thermodynamic properties of substances by means of surfaces. *Conn. Acad.*, II-2:382–404, 1873.
- [70] L. Girardin and D. Brodbeck. *Visual Approach for Monitoring Logs*. 12th Systems Adm. Conf., Boston, 1998.
- [71] A. Goel. *Vizcraft: A Multidimensional Visualization Tool for Aircraft Configuration Design*. IEEE Vis Conf., 425–8, 1999.
- [72] H. Goldstein. *Classical Mechanics*, 2nd ed. Addison-Wesley, Reading Mass., 1980.
- [73] E. Goursat. *A Course in Mathematical Analysis vol. I*. Ginn & Co, New York, 19xx.
- [74] M.J. Greenberg. *Euclidean and Non-Euclidean Geometries, Development and History*. W.H. Freeman and Co., 1974.
- [75] W.H. Guggenheimer. *Differential Geometry*. McGraw-Hill, New York, 1963.
- [76] H.S. Guyford and J.J. Haggerty. *Flight – Life Science Library*. Time Inc., New York, 1965.
- [77] F. Harary. *Graph Theory*. Addison-Wesley, Reading, Mass., 1969.
- [78] J. Harris. *Algebraic Geometry*. Springer-Verlag, New York, 1992.
- [79] H. Hauser. *Parallel Sets: Visual Analysis of Categorical Data*. Proc. IEEE Infovis, 2005.
- [80] M.A. Hearst. Clustering versus faceted categories for information exploration. *Comm. ACM*, 49–4:59–61, 2006.
- [81] N. Helfman. *Navigating in Visual Models of Decision Support Systems*. M.Sc. thesis, Dept. Comp. Sci., Bar Ilan University, Israel, 2000.
- [82] J. Helly. *Applications of Parallel Coordinates to Complex System Design and Operation*. Proc. Nat. Comp. Grap. Assoc. vol. III, 541–546, 1987.

- [83] P. Hertzog. *Visualizations to Improve Reactivity towards Security Incidents inside Corporate Networks*, SIGSAC Conf. Proc., 95–102. ACM, New York, 2006.
- [84] B. Hetzler, P. Whitney, L. Martucci and J. Thomas. *Multi-faceted Insight Through Interoperable Visual Information Analysis Paradigms in Proc. IEEE on Info Vis pp. 137–144*. IEEE Comp. Soc., Los Alamitos, CA, 1998.
- [85] D. Hilbert and S. Cohn-Vossen. *Geometry and the Imagination*. Chelsea, New York, 1983.
- [86] W. Hodge and D. Pedoe. *Methods of Algebraic Geometry, Vol. II*. Cambridge Univ. Press, Cambridge, 1952.
- [87] P. Hoffman and G. Grinstein. *Multidimensional Information Vis. for Data Mining with Applications for Machine Learning Classifications*. In Info. Vis. for KDD, Morgan-Kaufman, 2000.
- [88] P. Hoffman, G. Grinstein, K. Marx and I. Grosse. *DNA Visual and Analytic Data Mining*. Proc. IEEE Infovis, pp. 437–441, 1997.
- [89] Q. Huanin, C. Wing-Yi, X. Anbang, C. Kai-Lun, L. Kai-Hon and G. Ping. Visual analysis of the air pollution problem in hong kong. *Trans. on Vis. & Comp. Grap.*, 13–6:1408–15, 2007.
- [90] H.P. Hudson. *Ruler & Compass in ‘Squaring the Circle’ by Hopson et al*. Chelsea, New York, 1953.
- [91] M.Y. Huh and K.Y. Song. DAVIS: A java based data visualization system. *Comp. Stat.*, 17–3:411–423, 2002.
- [92] C.K. Hung and D. Ierardi. *Constructing Convex Hulls of Piecewise Smooth Objects*. Univ. of S. Calif. CS Dept. Tech R. 94–583, 1994.
- [93] C.K. Hung and A. Inselberg. *Parallel Coordinate Representation of Smooth Hypersurfaces*. USC Tech. Report #CS-92-531, Los Angeles, 1992.
- [94] C.K. Hung and A. Inselberg. *Description of Surfaces in Parallel Coordinates by Linked Planar Regions, Mathematics of Surfaces XII, 177–208, LNCS 4647*. Springer-Verlag, New York, 2007.
- [95] C.B. Hurley and R.W. Olford. Pairwise Display of High Dimensional Information via Eulerian Tours and Hamiltonian Decompositions. *to appear*, 2008.
- [96] IEEE. The faa’s advanced automation program — special issue. *IEEE Computer J.*, 20–2, 1987.
- [97] A. Inselberg. *N-Dimensional Graphics, LASC Tech. Rep. G320-2711, 140 pages*. IBM, 1981.

- [98] A. Inselberg. *Intelligent Instrumentation and Process Control, Proc. 2nd IEEE Conf. on AI Appl.*, 302–307. IEEE Comp. Soc., Los Alamitos, CA, 1985.
- [99] A. Inselberg. The plane with parallel coordinates. *Visual Computer*, 1:69–97, 1985.
- [100] A. Inselberg. *Parallel Coordinates: A Guide for the Perplexed*, in *Hot Topics Proc. of IEEE Conf. on Visualization*, 35–38. IEEE Comp. Soc., Los Alamitos, CA, 1996.
- [101] A. Inselberg. *Multidimensional Detective*, in *Proc. of IEEE Information Visualization '97*, 100–107. IEEE Comp. Soc., Los Alamitos, CA, 1997.
- [102] A. Inselberg. Visual data mining with parallel coordinates. *Comput. Statist.*, 13–1:47–64, 1998.
- [103] A. Inselberg and T. Avidan. *The Automated Multidimensional Detective*, in *Proc. of IEEE Information Visualization '99*, 112–119. IEEE Comp. Soc., Los Alamitos, CA, 1999.
- [104] A. Inselberg, M. Boz and B. Dimsdale. *Planar Conflict Resolution Algorithm for Air-Traffic Control and the One-Shot Problem*, in *IBM PASC Tech. Rep. G320-3559*. IBM Palo Alto Scientific Center, 1991.
- [105] A. Inselberg and B. Dimsdale. *Parallel Coordinates: A Tool For Visualizing Multidimensional Geometry*, in *Proc. of IEEE Conf. on Vis. '90*, 361–378. IEEE Comp. Soc., Los Alamitos, CA, 1990.
- [106] A. Inselberg and B. Dimsdale. Multidimensional lines i: Representation. *SIAM J. of Applied Math.*, 54–2:559–577, 1994.
- [107] A. Inselberg and B. Dimsdale. Multidimensional lines ii: Proximity and applications. *SIAM J. of Applied Math.*, 54–2:578–596, 1994.
- [108] A. Inselberg and C.K. Hung. *A New Representation of Hypersurfaces Based on Their Tangent Hyperplanes*. In *Course Notes for Tutorial #16 on Multidimensional Visualization* 105–112, SIGGRAPH94, ACM Publications, 1994.
- [109] A. Inselberg and P.L. Lai. *Visualizing Families of Close Planes*, 66. *Proc. 5th Asian Conf. on Stat.*, Hong Kong, 2005.
- [110] A. Inselberg, M. Reif and T. Chomut. Convexity algorithms in parallel coordinates. *J. ACM*, 34:765–801, 1987.
- [111] T. Itoh, et al. Hierarchical visualization of network intrusion detection data. *IEEE Comp. Graph. & Appl.*, 26–2:40–47, 2006.
- [112] M.Y. Ivory and M.A. Hearst. Statistical profiles of highly-rated web sites. *IEEE Inter. Comp.*, 6–2:55–63, 2002.

- [113] J. Johansson, P. Ljung, M. Jern and M. Cooper. *Revealing Structure within Clustered Parallel Coordinates Displays*, *Proc. IEEE Infovis*, 125–132. IEEE Comp. Soc., Los Alamitos, CA, 2005.
- [114] J. Johansson, P. Ljung, M. Jern and M. Cooper. Revealing structure in visualizations of dense 2d and 3d parallel coordinates. *Inf. Vis. Pal-Grave-Mcmillan Journ.*, 5:125–136, 2006.
- [115] C. Jones. *Visualization and Optimization*. Kluwer Academic Publishers, Boston, 1996.
- [116] M.W. Kantor, A.R. Liebler, E.S. Payne and E.E. Shult. (eds.) *Finite Geometries, Buildings, and Related Topics*. Clarendon Press Oxford, Oxford, 1990.
- [117] F. Karteszi. *Introduction to Finite Geometries*. North-Holland, Amsterdam-Oxford, 1976.
- [118] D.A. Keim. Designing pixel-oriented visualization techniques: Theory and applications. *Comp. Graphics & Applic.*, 6–1:59–78, 2000.
- [119] D.A. Keim and H.P. Kriegel. Visualization techniques for mining large databases: A comparison. *Trans. Knowl. and Data Engr.*, 8–6:923–938, 1996.
- [120] M. Kipouros, T. Mleczko and M. Savill. *Use of Parallel Coordinates for Post-Analyses of Multiobjective Aerodynamic Design Optimization in Turbomachinery (to appear)*. *Proc. 4th AIAA Meet.*, 2008.
- [121] M. Kline. *Mathematical Thought from Ancient to Modern Times*. Oxford Univ. Press, New York, 1972.
- [122] R. Kriz. *Visual Thinking by Scientists*. www.sv.vt.edu/classes/ESM4714/Gen-Prin/vizthink.html, 2007.
- [123] N. Kumasaka and R. Shibata. *High Dimensional Data Visualization the Textile Plot*. *Proc. Compstat 2006* Psysica-Verlag, 581–9, 2006.
- [124] P.L. Lai and C. Fyfe. A family of canonical correlation analysis networks. *Neural Proc. Let.*, 14–2:93–105, 2001.
- [125] F. Lauro, N.C. Palumbo and A. Iodice D’Enza. *New Graphical Symbolic Objects Representation in Parallel Coordinates, in Between Data Science and Applications (M. Schader et al. eds)*. Springer, Heidelberg, 2003.
- [126] C.L. Lay. *Linear Algebra and its Applications*. Updated Edition, Addison-Wesley, New York, 2000.
- [127] R.O. Lejeune. *Government Provided Complex Scenarios for the Advanced Automated System Design Competition Phase, MTR-85W240*. MITRE Co, McLean, Virginia, 1985.

- [128] M.M. Lipschutz. *Differential Geometry*. Schaum's Outlines on Math., McGraw-Hill, New York, 1969.
- [129] J. Martin and G. Kennedy. *Using Curves to Enhance Parallel Coordinates Visualization*. Proc. IEEE Infovis, London, IEEE Comp. Soc., Los Alamitos, CA, 10–16, 2003.
- [130] Mathcurve.com. *Great Website for Surfaces*. www.mathcurve.com/surfaces/surfaces.shtml, 2006.
- [131] T. Matskewich, A. Inselberg and M. Bercovier. *Approximated Planes in Parallel Coordinates*. Proc. of Geom. Model. Conf., St. Malo, Vanderbilt Univ. Press, 257–266, 2000.
- [132] B.H. McCormick, T.A. Defanti and M.D. Brown. *Visualization in Scientific Computing*. Computer Graphics 21–6, ACM SIGGRAPH, New York, 1987.
- [133] N. Megido. *On the Complexity of some Geometric Problems in Unbounded Dimension*, in *IBM Res. Rep. RJ5744(58033)*. IBM Research, 1987.
- [134] D. Michie, D.J. Spiegelhalter and C.C. Taylor. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood series in AI, 1994.
- [135] T. Mihalisin, J. Timlin and J. Schwegler. Visualizing multivariate functions, data and distributions. *IEEE Comp. Graph. and Appl.*, 1991.
- [136] T.M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [137] J.W. Moon. *Various Proofs of Cayley's Formula for Counting Trees, Seminar in Graph Theory, F. Harary (ed.)*. Holt, Rinehart & Winston, 1967.
- [138] J.R. Munkres. *Topology, 2nd. ed.* Prentice-Hall, New Jersey, 2000.
- [139] J. Nakano and K. Honda. *Three-dimensional parallel coordinates plot used for variable selection*. Proc. COMPSTAT 2006, (eds. A. Rizzi and M. Vichi), Springer, 187–195, 2006.
- [140] J.R. Newman. *World of Mathematics*. Simon and Schuster, New York, 1956.
- [141] M. d'Ocagne. *Coordonnees paralleles et axiale*. Gautier-Villars, Paris, 1885.
- [142] M. d'Ocagne. *Traite de Nomographie*. Gautier-Villars, Paris, 1899.
- [143] W.F. Osgood. *Advanced Calculus*. Macmillan, New York, 186–194, 1925.
- [144] E. Otto. *Nomography. Translated from 1956 Polish ed. by J. Smolska*. Macmillan, New York, 1963.
- [145] D. Pedoe. *Circles*. Pergamon, New York, 1957.
- [146] P.A. Penna and R.R. Patterson. *Projective Geometry and its Application to Computer Graphics*. Prentice-Hall, New Jersey, 1986.

- [147] Y. Pham, B. Cai and R. Brown. *Visualization Techniques for Tongue Analysis in Traditional Chinese Medicine*. Proc. Med. Imag. SPIE 14–19, 2004.
- [148] F.P. Preparata and M.I. Shamos. *Computational Geometry*. Springer-Verlag, New York, 1985.
- [149] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufman, 1993.
- [150] J. Riordan. *An Introduction to Combinatorial Analysis*. John Wiley, New York, 1958.
- [151] H. Rosenbaum and R. Schumann. *Chances and Limits of Progression in Visualization, in SimVis. Proc. Sim. & Vis.*, Magdeburg, Germany, 2007.
- [152] B.A. Rosenfeld. *A History of Non-Euclidean Geometry*. Springer-Verlag, New York, 1988.
- [153] F. Rossi. *Visual Data Mining and Machine Learning, in 14th Europ. ESANN Proc.*, 251–264. ESANN, Bruges Belgium, 2006.
- [154] P. Samuel. *Projective Geometry*. Springer-Verlag, New York, 1988.
- [155] M. Schall. Diamond and ice: Visual exploratory data analysis tools. *Perspective, J. of OAC at UCLA*, 18(2):15–24, 1994.
- [156] C. Schmid and H. Hinterberger. *Comparative Multivariate Vis. Across Conceptually Different Graphic Displays, in Proc. of 7th SSDBM*. IEEE Comp. Soc., Los Alamitos, CA, 1994.
- [157] O. Schreier and Sperner. *Projective Geometry of n Dimensions*. Chelsea Publishing Company, New York, 1985.
- [158] N. Shang and L. Breiman. *Distribution based trees are more accurate*. Proc. of 1st Inter. Conf. on Neural Info. Proc. (ICONIP96), 133, 1996.
- [159] B. Shneiderman. *Leonardo's Laptop*. MIT Press, Boston, 2002.
- [160] B. Shneiderman and C. Plaisant. *Designing the User Interface (4th Edition)*. Addison Wesley, Reading MA, 2005.
- [161] C.H. Sisam and W.F. Atchison. *Analytic Geometry, third ed.* Holt, Rinehart and Winston, New York, 1955.
- [162] D.J. Sobiski. *Collision Detection and Collision Avoidance, Document Nos. 196A798*. LORAL Federal Systems Co, Oswego, NY, 1995.
- [163] D.M.Y. Sommerville. *An Introduction to the Geometry of N Dimensions*. (First publ. 1929), Dover Publications, New York, 1958.
- [164] R. Spence. *Information Visualization: Design for Interaction, 2nd ed.* Prentice Hall, 2007.
- [165] I. Stewart. *Flatterland*. Perseus, New York, 2001.

- [166] D.J. Struik. *Lectures on Classical Differential Geometry 2nd. Ed.* Dover (Addison-Wesley 1961), New York, 1988.
- [167] D.F. Swayne, D. Cook and A. Buja. *XGobi: Interactive Dynamic Graphics in the X Window System*. JCGS, 7–1, 113–130, 1998.
- [168] D.F. Swayne, (T. Lang) Duncan, A. Buja and D. Cook. Ggobi: evolving from Xgobi into an extensible framework for interactive data visualization. *J. of Comp. & Graph. Stat.*, 43:423–444, 2003.
- [169] J.L. Synge and A. Schild. *Tensor Calculus, 2nd ed.* University of Toronto Press, Toronto, 1956.
- [170] V. Tontodonato. *Symbolic Objects Representation, Ph.D. thesis.* Dept. of Math and Stat., Univ. Federico II, Napoli, 2002.
- [171] M. Tory, S. Potts and T. Moller. A parallel coordinates style interface for exploratory volume visualization. *Trans. on Vis. & Comp. Grap.*, 11–1: 71–80, 2005.
- [172] E.R. Tufte. *The Visual Display of Quantitative Information*. Graphic Press, Connecticut, 1983.
- [173] E.R. Tufte. *Envisioning Information*. Graphic Press, Connecticut, 1990.
- [174] E.R. Tufte. *Visual Explanation*. Graphic Press, Connecticut, 1996.
- [175] L.A. Tweedie, R. Spence, H. Dawkes and Su H. *Externalizing Abstract Mathematical Models, Proc. CHI, Vancouver, Can.*, 406–412. ACM Press, 1996.
- [176] A.R. Unwin. Requirements for interactive graphics software for exploratory data analysis. *Comp. Stat.*, 14:7–22, 1999.
- [177] A.R. Unwin, M. Theus and H. Hofmann. (Eds.) *Graphics of Large Datasets*. Springer, New York, 2006.
- [178] F.B. Viegas, M. Wattenberg and D. Kushal. *Studying Cooperation and Conflict between Authors with historyflo Visualizations in Proc. of SIGHCI on Human Factors in Computing Systems*, 575–582. ACM Press, New York, 2004.
- [179] R.J. Walker. *Algebraic Curves*. Springer, New York, 1978.
- [180] W.A. Walker. What is a point of inflection? *Amer. Math. Monthly*, 63: 182–183, 1956.
- [181] M.O. Ward. *XmdvTool: Integrating multiple methods for visualizing multivariate data, Proc. IEEE Conf. on Vis., CA*, 326–333. IEEE Comp. Soc., Los Alamitos, CA, 1994.
- [182] C. Ware. *Information Visualization: Perception and Design, 2nd ed.* Morgan Kaufman, San Francisco, 2004.

- [183] C.E. Weatherburn. *Differential Geometry of Three Dimensions*. Cambridge Univ. Press, Cambridge, 1955.
- [184] E. Wegman. Hyperdimensional data analysis using parallel coordinates. *J. Amer. Stat. Assoc.*, 85:664–675, 1990.
- [185] P.C. Wong and R.D. Bergeron. *30 Years of Multidimensional Multivariate Visualization in Scientific Visualization: Overviews, Methodologies & Techniques*, G.M. Nelson, H. Mueller and H. Hagen (eds.), 3–33. IEEE Comp. Soc., Los Alamitos, CA, 1997.
- [186] V.O. Ya. Some integral calculus based on euler characteristic, lect. notes. *Amer. Math. Monthly*, 1346:127–138, 1988.
- [187] L. Yang. Pruning & visualizing generalized association rules in parallel coordinates. *IEEE Know. & Data Engr.*, 15(1):60–70, 2005.
- [188] H. Ye and Z. Lin. Speed-up simulated annealing by parallel coordinates. *Euro J. of OR*, 173:59–71, 2006.
- [189] X. Yin, et al. *VisFlowConnect*. Comp. Sec., ACM Workshop on Visualization, Washington, D.C. 26–34, 2004.

Index

- Abbot A.E., 2
- Adjashvili D., xxiii, 376
- air traffic control
 - angle between superplanes, 133
 - collision avoidance, 103–106
 - conflict intervals, 106
 - conflict resolution, 108–113
 - information display, 101
 - particle fields, 103–113
- angle between two planes, 133
- approximate quadric hypersurfaces, 359–365
- Archimedes, xv
- astroidal surface, 370
- Atariah, D., 322
- automatic classification, 406
- Avidan S. and T., xxiii
- axes
 - axis translation \rightarrow rotation, 133
 - permutations, 122
 - spacing, 125

- Banchoff T.F., 2
- Bertin J., xxv
- Binniaminy M., 322
- Boz M., xxiii
- bump, 370

- Chatterjee A., xxiii, 250, 326
- Chen C., xxv
- Chen C.C., xxv
- Cohen L., xxiii, 376
- Cohen S., 429
- Cohen-Ganor S., 85
 - displaying several lines, 429–442
- collinearity construction algorithm
 - example in \mathbb{R}^6 , 173
- collinearity construction algorithm \mathbb{R}^N , 170
- collinearity property p -flats in \mathbb{R}^N , 169

- computer graphics
 - transformations and matrices, 31–35
- conoids, 352
 - parabolic conoid, 352
 - Plücker's conoid, 352
 - Whitney's umbrella, 352
- construction algorithms
 - in \mathbb{R}^3 , 136–159
 - half spaces, 136
 - intersecting planes, 138
 - line on a plane, 136
 - parallel plane on point, 151
 - plane line intersection, 145
 - rotation \leftrightarrow translation, 151
 - special planes, 141
 - the four indexed points, 138
 - in \mathbb{R}^4 , 174–179
 - intersecting hyperplanes, 175
 - the five indexed points, 174
- convex surfaces, 367
- course
 - prerequisites, xxii
 - projects, xix
 - syllabus, xviii–xxii
- curves, 195–241
 - algebraic curves, 208
 - Plücker class formula*, 230
 - Plücker curves*, 229–230
 - examples, 208–211
- conics
 - classification of transforms, 220–226
 - ideal points, 226–227
 - the Indicator, 226
 - transforms via Möbius transformation, 217
- conics \leftrightarrow conics, 216–229
- convex sets and their relatives, 231–241
- convex upward and downward, 197
- curve plotting, 216
- cusps and inflection points, 199–204

- duality
 - bitangent* \leftrightarrow *double-point*, 230
 - envelope, 43
 - exponential, 212
 - gconics – generalized conics, 231–241
 - convex hull with curved edges, 235
 - convex-hull construction, 235
 - gconics \leftrightarrow gconics, 232–241
 - inner intersection, 238
 - operational dualities, 238
 - outer union, 238
 - outer union \leftrightarrow inner intersection, 240
 - Inselberg transformation, 206
 - line, 43
 - oscillatory, 212
 - plotting, 207
 - point, 43
 - point-curves \leftrightarrow point-curves, 204–241
 - point-curves and line-curves, 195–196, 204
 - space curves, 216
 - trigonometric, 212
- d'Ocagne, xvi
- data mining, xvi, xvii, xxi, 378–427
 - classification, xvii, xxi
 - grand vision, 5
 - guidelines and strategies, 4
 - introduction, 379–382
 - Piatetsky-Shapiro G., xxv
 - Zaavi J., xxv
- decision support, xxi, 359, 416–418
- Desargue
 - theorem, 10
- developable surfaces
 - characteristic point, 312
 - duality with space curves, 312
 - general cone, 341
 - helicoid, 342
 - reconstruction, 319
 - reconstruction from its rulings, 321
 - representation, 314
 - ruling, 312
 - rulings of cylinders, 326
 - start points, 319
- symmetry and coincidence, 333
- tangent planes, 311
- dimensionality, 2
- dimple, 370
- Dimsdale B., xxiii, 361, 375
 - conics, 217
 - hyperplanes, 165
- duality, 10
 - \mathbb{R}^2 – slope and x -coordinate relation, 51
 - bitangent* \leftrightarrow *double point*, 230
 - cuspidal \leftrightarrow inflection point, 199–204
 - gconics – generalized conics
 - operational dualities, 238
 - outer union \leftrightarrow inner intersection, 240–241
 - in 2-D point \leftrightarrow line, 15
 - in 3-D point \leftrightarrow plane, 15
 - point \leftrightarrow line in \mathbb{P}^2 , 49–62
 - point \leftrightarrow line
 - horizontal position of $\bar{\ell}$, 51
 - line \rightarrow point linear transformation, 53
 - point \rightarrow line linear transformation, 54
 - point-curves and line-curves, 55
 - projective plane, 49
 - recognizing orthogonality, 58
 - rotation \leftrightarrow translations, 57
 - transformations, 56–60
- Ehrenfest P., 2
- Eickemeyer
 - representation of p -flats, 167
 - representation of hyperplanes, 166
 - superplanes – \mathcal{E} , 125
- Eickemeyer J., xxiii, 125
- ellipsoid, 367
- envelopes, 55, 118, 185–194
 - formulation and computation, 186
 - necessary and sufficient conditions, 188
 - of families of curves, 190–193
 - one-parameter family of curves, 185
 - singular points, 189

- Euclid
 - elements, 8
 - parallels axiom, 15
- exploratory data analysis – EDA, xvii, xxi, 382–383
 - data display, 380
 - hundreds of variables, 400
 - interactivity, 380
 - pattern recognition, 379
 - simple case study, 383–390
 - software, 4
 - software requirements, 4
 - testing assumptions, 401
- exploratory data analysis – EDA, xxiii
- exterior point, 349
- financial dataset, 392, 401
- Fiorini P., xxiii
- geometry, 7–48
 - affine, 48
 - constructions
 - compass alone, 8
 - first algorithms, 8
 - Poncelet and Steiner conditions, 8
 - Egyptians, 7
 - Euclid
 - elements, 8
 - Greeks, 7
 - origins, 7
 - parallelism, 2
 - projective, 8–45
 - transformations
 - invariants, 45–48
- geovisualization, xxiv
- Gibbs W.J., 333
- Giles N., 236
- GIS, xxiv
- gold, high price
 - correlation with currencies, 397
- grand tour
 - Asimov D., xxiv
- Greenshpan O., 460–475
- Grinstein G., xxv
- group, 46
- half plane
 - oriented, 197
- half spaces, 136
- Hearst M., xxv
- Helfman N.
 - decision support, xxiii
- helicoid (developable), 342
- Hinterberger H.
 - comparative visualization, xxiv
 - data density analysis, xxiv
- homogeneous coordinates, 22–43
 - analytical proofs, 39–43
 - for finite geometries, 24
 - on a line, 25
- Hung C.K., xxiii, 314, 326, 335, 336, 375
 - convex-hull with curved edges, 235
 - surface representation decomposition, 451–460
- Hurwitz A., xxiii
- hypercube, 297
- hyperellipse, 362
- hyperellipsoids, 360
- ILM
 - duality in \mathbb{P}^2
 - point \leftrightarrow line, 51, 56, 58
 - geometry
 - Desargues’s theorem, 42
 - Pappus’s theorem, 42
 - point \leftrightarrow line duality, 19
 - multidimensional lines
 - general case, 83
 - indexed points – adjacent and base variable, 67
 - three-point collinearity, 71
 - planes
 - rotation \leftrightarrow translation for indexed points, 122, 153
- incidence matrix, 300
- intelligent process control, 363–365
 - domain, 363
 - instrument panel, 364
 - learning, 365
 - rule, 363
- interaxis distances
 - recursive notation, 299
- Itai T., 322

- Keim D., xxiv
- Kosloff Z., 322
- Kriz R., 2

- Lifshits Y., 373

- Möbius strip, 354
- Makarov I., 373
- Matskewich T., xxiii, 250, 273, 279
 - proximate hyperplanes construction, 279
 - proximate hyperplanes representation, 273
- multidimensional lines
 - 3-point collinearity, 127
 - air traffic control, 100–113
 - construction algorithms, 79–82
 - example, 80
 - displaying several lines, 429–442
 - distance and proximity, 88–100
 - L_2 minimum distance – monotonicity with dimension, 97–100
 - L_2 versus L_1 minimum distance, 99
 - constrained L_1 distance, 92–96
 - intersecting lines, 88
 - minimum distance between lines, 91–100
 - nonintersection, 89–100
- indexed points
 - 3-point collinearity, 70–72
 - arbitrary parametrization, 75
 - dummy index, 77
 - identification, 82
- representation, 63–88
 - adjacent variables, 63–65
 - base variable, 65
 - convenient display of a line with positive slopes, 83
 - convenient display of several lines, 85
 - general case, 75–78
 - line on a plane, 119
 - transforming positive to negative slopes, 83–85
 - two-point form, 68–70
 - two-point form continued, 78
- rotations \leftrightarrow translations duality, 87
- separation in the xy plane
 - above and below relations, 149

- nearly developable surfaces, 343
- network
 - visualization and analysis, 460
- nomography, xvi
- nonlinear models, 416–418
- nonorientability
 - in \mathbb{R}^2 , 149
 - projective plane \mathbb{P}^2 , 13
 - visualization, 358
- nonorientable surface, 354

- paraboloid, 367
- parallel coordinates
 - \mathbb{P}^2 , 49–62
 - bitangent* \leftrightarrow *double-point* duality, 230
 - air traffic control, 100–113
 - algebraic curves, 229–230
 - applications, xxiv, 5
 - automatic classification, 406
 - automatic classifiers, 4
 - bare essentials, 427
 - complex-valued functions, 476–488
 - conics, 216–229
 - curves, 55, 195–241
 - curve plotting, 207–216
 - cusp \leftrightarrow inflection-point duality, 199–204
 - data mining, 378–427
 - decision support, 416–418
 - definition, 3
 - enhancing with curves, xxiv
 - exploratory data analysis – EDA, 382–383
 - atomic queries, 385
 - guidelines, 383–390
 - multidimensional detective, 382–383
 - requirements, 381
 - first proposed, 2
 - gconics – generalized conics
 - operational dualities, 238

- hypercube, 61
- in 3-D, xxiv
- Inselberg transformation, 206
- multidimensional lines, 63–113
 - representation, 70
- network visualization and analysis, 460–475
- notation in \mathbb{R}^2
 - bar, point, line, 55
- orthogonality, 58
- parallax, 4
- parallelism, 51
- permutations for adjacencies, 388
- planes, p -flats, and hyperplanes, 115–183
- point \leftrightarrow line duality
 - queries, 387
- properties, 3
- proximate planes and hyperplanes, 243–295
- recursive construction algorithm, 4
- representation mapping I, 73–75
- separation in \mathbb{R}^2 and \mathbb{R}^3 , 149
- separation in the xy plane, 196
- software, 4
- starplots, xxiv
- surfaces in \mathbb{R}^N , 297–377
- visual data mining – EDA, 4
- Pascal, 44
- pattern recognition, 1
- pi (π), 7
- planes, p -flats, and hyperplanes
 - approximate planes and flats, 179
 - axis spacing, 125
 - axis translation \rightarrow rotation
 - rotation angle in \mathbb{R}^3 , 133
 - coefficients and distance between adjacent points, 134
 - collinearity construction algorithm
 - example in \mathbb{R}^6 , 173
 - collinearity construction algorithm \mathbb{R}^N , 170
 - coplanarity
 - in industrial data, 117
 - indexed points, 125
 - near coplanarity, 179
 - vertical lines pattern, 116
 - determining multiple planes, 180
 - line on a plane, 119
 - planes in \mathbb{R}^3 , 115
 - planes on a line, 122
 - point on a plane, 118
 - recursive construction, 164
 - recursive construction algorithm \mathbb{R}^N , 170
 - representation
 - p -flat in \mathbb{R}^N , 167
 - augmented, 132
 - by two indexed points in \mathbb{R}^3 , 128
 - collinearity property for p -flats, 169
 - hyperplane in \mathbb{R}^N , 166
 - hyperplanes and p -flats in \mathbb{R}^4 , \mathbb{R}^5 , 168
 - indexed points, 124
 - indexed points in \mathbb{R}^N , 164
 - planar coordinates, 118
 - recursive construction, 164
 - recursive construction algorithm, 127
 - unification for p -flats $0 \leq p < N$, 168
 - vertical lines, 115
- rotation \leftrightarrow translation indexed points, 151
- rotation \leftrightarrow translation vertical lines, 120
- separating points clusters on different planes, 443–451
- separation in \mathbb{R}^3 – points and planes, 149
- superplanes sp , 125
 - in \mathbb{R}^4 , 159
- upper-planes sp
 - in \mathbb{R}^3 , 126
- point \leftrightarrow line duality in \mathbb{P}^2 , 49–62
- projective geometry, 9–45
 - 3-point collinearity, 24
 - analytic, 22–43
 - proofs, 39–43
 - augmented lines, planes, space, 12
 - axioms, 10, 13–16
 - conics, 43
 - cross-ratio, 27
 - invariant under perspectivity, 27
 - Desargue’s theorem, 12
 - duality, 9, 16–19

- finite geometries, 19–21
- fundamental transformations, 26–29
- homogeneous coordinates, 9, 22–43
 - analytical proofs, 39–43
 - for finite geometries, 24
 - on a line, 25
- ideal elements, 11
- incidence relation, 10
- line coordinates, 23
- linearizability of projective transformations, 29–31
- Pappus's theorem and its dual, 18
- Pappus's theorem generalizations, 44
 - Brianchon's theorem dual of Pascal's theorem, 45
 - Pascal theorem for conics, 44
- perspectivity and projectivity, 28
- perspectivity, 26
- primitives, 10
- projections and sections, 10
- projective plane model, 10–13
 - Hilbert, 12
- projective plane nonorientability, 13
- projective transformations
 - in \mathbb{P}^2 , 32
 - in \mathbb{P}^3 , 33
 - linearized, 31–35
 - perspective, 32, 34
 - rotation, 32, 33
 - scaling, 32
 - translation, 32, 33
- proximate planes and hyperplanes, 243–295
 - Chatterjee, A., 250
 - Matskewich, T., 250
 - neighborhood, 244
 - outline of development, 253
- proximity of hyperplanes, 251–295
 - adding the c_0 , 280–285
 - components and structure of Ω , 257–263
 - conditions for Ω being bc , uc , or gh , 267–274
 - construction of Ω , 263–275
 - formulation of the problem, 251–254
 - full image of \overline{NH} , 275–279
 - matching points and navigation, 287–295
 - properties of f_N , 254–257
 - region Ω , 257–275
 - region crossing x axis, 287
 - some examples, 283–295
- proximity of lines and line neighborhoods, 244–251
- topology for proximity, 243–245
- quasars data set, 390–392
- queries
 - atomic, 385
 - compound, 401
 - zebra, 397
- recursive construction algorithm, 127
 - general in \mathbb{R}^N , 164
- Reif M., xxiii
- representation
 - astroidal representation, 370
 - bump, 370
 - convex surfaces, 369
 - dimple, 370
 - Möbius strip, 358
 - sphere, 367
 - torus, 370
- representation mapping
 - version II, 180
- Rivero J., xxiii
- rotation, 158
- rotation \leftrightarrow translation points, 151–158
- rotation \leftrightarrow translation vertical lines, 120
- rotations
 - coordinate system in \mathbb{R}^3 , 36
 - in \mathbb{R}^2 , 32
 - in \mathbb{R}^3 , 33
- ruled surfaces
 - conoids, 352
 - Möbius strip, 354
 - representation, 349
 - saddle, 347
- saddle, 347
- Satellite data set, 383–390
- separation in \mathbb{R}^2 – points and lines, 149

- separation in \mathbb{R}^3 – points and planes (new approach), 146
- separation in \mathbb{R}^3 – points and planes (old approach), 149
- separation in the xy plane
 - above and below relations, 196
 - convex upward or downward, 197
 - oriented half plane, 196
- Shahaf N., 443–451
- Shneiderman B., xxv
- Singer Y., 460–475
- Snow J., xv
- Spence R., xxv
- sphere, 367
- Spivak M., 376
- surfaces
 - representation decomposition and developable quadrics, 451–460
- surfaces in \mathbb{R}^N , 297–377
 - approximate quadrics, 359–365
 - decision support, 359
 - hyperellipsoids, 360
 - intelligent process control, 362–365
 - boundary contours, 304–310
 - developable surfaces, 310–346
 - ambiguities and uniqueness, 316–319
 - characteristic point, 312
 - duality with space curves, 312
 - general, 342–346
 - general cone, 341
 - helicoid, 342
 - Hung C.K., 311
 - reconstruction, 319–325, 341
 - reconstruction from its rulings, 321
 - representation, 314
 - ruling, 312
 - rulings of cylinders, 326
 - specific developables, 325–341
 - specific developables, cones, 336–341
 - specific developables, cylinders, 326–336
 - symmetry and coincidence, 333
 - tangent planes, 311
- formulation, 301–304
 - general
 - matching algorithm, 372–377
 - hypercube, 297
 - matching algorithm, 301, 303, 311, 348, 351, 352, 372
 - more general surfaces, 365–377
 - nearly developable surfaces, 343
 - normal vector, 302
 - preliminaries, 297–300
 - representation, 301
 - algebraic surface, 306
 - astroidal surface, 370
 - boundaries, 305
 - boundaries - example, 307
 - boundaries - the class \mathcal{EG} , 305
 - boundaries of quadrics, 306
 - boundaries of ruled surfaces, 348, 351
 - boundaries, ideal points, 305
 - convex surface boundaries, 366
 - convex surfaces, 367, 369
 - coordinates of representing points, 303
 - Möbius strip, 358
 - ruled surfaces, 347
 - saddle, 347
 - sphere, 367
 - torus, 370
 - representation of developable and ruled, 358
 - representing points relations, 300
 - ruled surfaces, 346–359
 - conoids, 352
 - Möbius strip, 354
 - representation, 349
 - saddle, 347
- to see \mathbb{C}^2 , 476–488
- torus, 370
- transformations
 - invariants, 45
 - linear, 46
 - linear fractional, 30
 - linearizability of projective transformations, 31
- Möbius, 30
 - homomorphic to group of nonsingular matrices, 217

- visualization
 - Archimedes, xv
 - early success of, xv
 - emergence of, xv
 - information, 1
 - multidimensional, xvi, 1
 - science, 2, 333
 - scientific, 1
- VLSI data set, 401–406
- Ward M., xxv
- Ware C., xxv
- Yaari Y., 476–488
- zero defects, 401–405